OMEGA

The International Journal of Management Science

Volume 4 1976

Author and Title Index

• **Chief Editor**

Professor SAMUEL EILON, DSc(Eng.), PhD — Department of Management Science, Imperial College, London, SW7 2BX

• **Associate Editors**

RJ BETTS and RB FLAVELL — Department of Management Science, Imperial College, London, SW7 2BX

Editorial Board

Professor RJ BALL, London Graduate School of Business Studies, Sussex Place, Regent's Park, London NW1 4SA

Professor BEN BERNHOLTZ, Department of Industrial Engineering, University of Toronto, Toronto 5

Professor C WEST CHURCHMAN, Space Sciences Laboratory, University of California, Berkeley, CA 94720

Professor BELA GOLD, Department of Economics, Case Western Reserve University, Cleveland, OH 44106

Professor CH GUDNASON, Danmarks Tekniske Højskole, København

Professor DOUGLAS C HAGUE, Manchester Business School, Booth Street West, Manchester M15 6PB

DAVID B HERTZ, McKinsey and Company Inc., 245 Park Avenue, New York, NY 10017

Professor KNUT HOLT, The Norwegian Institute of Technology, 7034 Trondheim

Professor HAROLD J LEAVITT, Graduate, School of Business, Stanford University, Stanford, CA 94305

Professor GERALD J LIEBERMAN, Department of Operations Research, Stanford University, Stanford, CA 94305

Sir FRANK MCFADZEAN, British Airways, PO Box 13, Air Terminal, Buckingham Palace Road, London SW1W 9SR

Professor DANIEL TEICHROEW, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48104

Pergamon Press Oxford · New York · Paris · Frankfurt

AUTHOR INDEX

Albu A 513
Allen TH 736
Alperovitch Y 358
Anvari M 225
Araten M 279
Ashton RH 609
Axelrod CW 321
Baligh HH 583
Banbury J 13
Bender AD 67
Berger PD 230
Berhold M 487
Berhold MH 165
Booth GG 85
Borcherding K 125
Borovits I 313
Bowman JP 85
Brooke MZ 149
Browne WG 265
Bunn DW 97
Burgoyne JG 13
Burton RM 583
Calpine HC 141
Chapman P 19
Chesterton BK 507
Conrad SA 93
Cosmetatos GP 100, 232, 489
Cottam M 205
Creighton JW 341
Damon WW 673
Dickman B 279
Dökmeci VF 617
Dorrell MG 149
Douglas B 67
Dyson RG 479
Eilon S 1, 119, 245, 350, 367, 501, 623
Ein-Dor P 313
Elster RS 341
Ezra D 629
Flavell RB 731
Forgionne GA 103, 228
Friedman Y 358
Geoffrion AM 49
Ghymn K II 595
Giles P 377
Gillett BE 711
Githens WH 341
Glazebrook KD 79
Gleason JM 353
Glover L 571
Golding A 141
Goodeve C 381
Gray DA 685
Gregory G 509, 643
Hadar J 181
Haines GH, Jr 253
Hannan EL 733
Hansen P 347
Harrison FL 447
Hefley GL 343
Hindelang TJ 738
Hindle D 491
Hindmarsh GW 234
Hitchings GG 205
Hodges SD 699
Honeyman N 187
Howarth JG 12
Hughes BP 12
Johnson JG 711
Kahalas H 685
King WR 9, 595
Kira D 355
Knolmayer G 493
Kosten L 345
Lamar C 253
Lethbridge DG 397
Lilly CC 353
Littauer SB 35
Locander WB 741
Loughridge DW 673
Martell TF 407
Meier P 378
Miles R 397
Mojena R 85
Moore CL 301
Moore JM 193
Moores B 59
Moskowitz H 125, 743
Moya-Angeler J 719
Musgrove GL 341
Nadler G 383
Oliver RL 455
Pavitt K 375, 539
Perrott BE 15
Peterson DW 673
Potts JWF 187
Pye R 289
Pyle EB, III 67
Radford KJ 559
Randolph PH 463
Reeves GR 339
Rothwell R 375
San Miguel JG 577
Scannell RW 741
Schaefer RE 125
Sergeant R 491
Sichel HS 417
Sieff M 633
Simon LS 253
Simpson TD 103
Sissouras AA 59
Sprague RH, Jr 657
Stainton RS 747
Starr MK 251, 630
Stephenson E 437
Steuer RE 455
Swaithes G 479
Tangedahl L 571
Taylor AJ 175
Theeuwes JAM 107
Tilanus CB 105, 107
Tomlinson RC 527
Towill DR 234
Varde VS 331
Wainstein BM 417
Watson HJ 657
West EN 225
Westlake WJ 67
Whitmore GA 215
Wiginton JC 109
Yegulalp TM 35
Zahariev GK 35
Zoltners AA 301

TITLE INDEX

Adaptive Trading Rules for Commodity Futures 407
Alternative Estimators of Lines of Securities 109
Analysis in Support of Strategic Policy Making 559
Application of Multiple Objective Linear Programming to Media Selection 455
Approach to Stochastic Programming for Medium Term Planning 479
Approximation for the Replenishment of Perishables Problem 489
Assessing Operational Policies 437
Bayesian Point Estimation on the Bernoulli Parameter 97
Cash Flow Models: A Review 643
Causes of the Decline in British Merchant Ship-Building and Marine Engineering 513
Comment on "A Dynamic Model of Process and Product Innovation" (Feedback) 375
Comment on "Stochastic Dominance for Ranking Ventures" 355
Communication of Control in an International Company: Methods of Investigation 149
Contractual Sharing Rules 487

Vol. 4
1976

Coordination of Short-Run Decision-Making with Long-Range Planning 673
Decision Support System for Banks 657
Design of a Strategic Planning Management Information System 595
Deterministic Model for the Housing Decision 85
Dynamic Strategies for Branch and Bound 571
Editorials 1, 119, 245, 367, 501, 623
Effect of Telecommunications on the Location of Office Employment 289
Effective Use of Computer Resources 321
Efficient Heuristic Procedure for Solving the Layout Design Problem 205
Energy Analysis: A Review of Methods and Applications 19
Equal Pay (Editorial) 245
Estimating Learning Time for Repetitive Tasks 234
Facilities Design with Graph Theory and Strings 193
Feedback 9, 251, 375, 507, 629
Fixed Costs Constraints and Realities 747
Food Preparation by Critical Path Method 230
Framework for Optimizing Managerial Decision 35
Government Policies Towards Innovation: A Review of Empirical Findings 539
Graphical Interpretation in Goal Programming Problem 733
Guided Tour of Recent Practical Advances in Integer Linear Programming 49
Heuristic for Sampling Rare Populations 225
How Many-Sided are Shadow Prices at Degenerate Primal Optima? 493
How Society Works (Editorial) 1, (Feedback) 381
How to Charge for Management Services (Editorial) 623
Human Relations in Industry 633
Implications of Goal Programming for Insurance Agency Decision Making 353
In-Company Training (Editorial) 119
In-Company Training—A Comment (Feedback) 629
Information Processing in Managerial Decision Making: A Preliminary Study 577
International Register of Theses and Dissertations 111, 237, 361, 495, 749
Inventory Decisions in Product-Mix Models 339
'Irrationality' of Managerial Judgments: Implications for Information Systems 125
Job Shop Scheduling—A Case Study 463
Linear Programming and Financial Cost Models for Industrial Enterprises 279
Macro-Efficiency in Interactive Work Units 103
Management Applications of the Inverse Gaussian Distribution 215
Management Careers for Management Scientists—Comments (Feedback) 12
Management Decision Rules in the Marketing of Electricity 228
Managerial Authority in a Decision Analysis Context 165
Managers' Uses of Models 253
Memoranda 97, 225, 339, 487, 731
Methodological Optimality in Operations Research (Feedback) 9
Model with Shortage of Places for Educational and Manpower Systems 719
Modelling of Supply in the Shipping Industry 175
Modelling the Product Elimination Decision 741
Multiproportional RAS Method 105
Multi-Terminal Vehicle-Dispatch Algorithm 711
Myth of Self-Actualization—A Comment (Feedback) 251
New Goal Programming Formulation 731
Note on "A Linear Programming Model for Determining and Optimal Regional Distribution of Petroleum Products" (Feedback) 378
Optimal Regional Distribution of Petroleum Products 301
'Optimum' Number of Beds in a Coronary Care Unit 59
Optimization Models for Fraction Defective Control Charts—A Markovian Approach 358
OR, Organisational Design and Adaptivity 527
Organization Structure and Cooperative Market Relations 583
Overflow Problems in File Organization 345
Paired Comparisons in Measuring Expected Benefit 491
Parkinson, Peter and Probability: Implications for Productivity 383
Pathways to Management Numeracy 397
Personnel Development and Evaluation 341
Planning Ambulatory Health Care Delivery Systems 617
Planning and Freedom (Editorial) 501
Planning for Teaching-Resource Requirements in Colleges of Technology 187
Policy Lapsation Model 331, (Feedback) 377
Price-Quality Relationship: A Heuristic Model 417
Problems in the Application of Portfolio Selection Models 699

Production Planning in Practice 447
Profitability Index for Alternative Research Projects 79
Quality Control Strategy 738
Quality of Working Life (Editorial) 367
Quantitative Model for Manpower Decision Making 685
Responsibilities of the Management Scientist (Feedback) 507, 630
Robustness of Linear Models for Decision-Making 609
Robustness of Linear Models for Decision-Making: Some Comments 743
Sampling Information in New Product Marketing 93
Scheduled and Charter Impairment in North Atlantic Air Services 265
Sensitivity of the Rate of Return 350
Simple Plant Location Problem 347
Simulation of R&D Investment Strategies 67
Some Numerical Equilibrium Results for the D/G/1:(∞/FIFO) Queueing Model 100
Some Numerical Equilibrium Results for the GI/D/1:(∞/FIFO) Queueing Model 232
Some Properties of Pareto-optimal Choices in Decision Problems 141
Stochastic Dominance for Ranking Ventures: Comments and Extensions 181
Two Issues in the Evaluation of Job-Shop Schedules 313
Using the Delta Chart in Project Management 736
Value Index Concept for Demand Forecasting 343
Variance Analysis Revisited 107

Annual Subscription Rates (1977)

Published bi-monthly. For libraries, university departments, government laboratories, industrial and other multiple-reader institutions U.S. $93.00 per annum (including postage and insurance). *Specially reduced rates to individuals:* In the interests of maximising the dissemination of material published in this important international journal we have established a two-tier price structure. Any individual, whose institution takes out a library subscription, may purchase a second or additional subscription for their use at the much reduced rate of $30.00 (£12.50) *per annum*. All subscription enquiries should be addressed to: The Subscription Fulfillment Manager, Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England.

Microform Subscription and Back Issues

Back issues of all previously published volumes are available in the regular editions and on microfilm and microfiche. Current subscriptions are available on microfiche simultaneously with the paper edition and on microfilm on completion of the annual index at the end of the subscription year.

Publishing, Subscription and Advertising Offices

Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England.
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford,
New York 10523, U.S.A.

Copyright 1976 Pergamon Press Ltd.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from the publishers. (See also note on copyright on inside back cover).

Printed in Great Britain by A. Wheaton & Co., Exeter

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

EDITORIAL

How Society Works

"IN ANY SOCIETY, individuals and social groups are bound to conflict over goals, priorities and their share of available resources. But they can still live tolerably well together if they accept certain minimal rules and principles which inescapably govern how their society functions." This is how Sir Charles Goodeve introduces a booklet entitled "How Society Works," recently published by OPUS, an Organization for Promoting the Understanding of Society (1975, London). The purpose of the booklet is to enunciate some general principles that apply to society—any society—so that its workings will thereby become better understood (the booklet is even subtitled "The need for wider understanding"). Eleven principles are proposed:

"1. Goods and services cannot be used or consumed unless they have been produced and made available.
2. Except in times of surplus, an increase in any one person's or group's consumption, if not matched by increased production, reduces the goods and services available for consumption by all other people in the same distribution area by the same total amount.
3. Of the total effort and resources available, a higher proportion can be devoted to investment only if a lower proportion is devoted to consumption.
4. Prices cannot be held below costs.
5. Man has the capacity to take in information, assess this in relation to his previous learning and choose a line of action in the light of his recognition of the related restraints and pressures. His choice usually indicates what he thinks will bring an outcome most advantageous to himself.
6. Membership of a group is both a means through which the individual achieves his objectives and also a constraint on his choices, in that he experiences pressures towards conformity to group standards and towards co-operation with his fellows.
7. Individuals, groups and organisations tend to cope with conflict by denying responsibility for their actions and passing on the responsibility to others.

8. All known systems of distribution of goods and services and of jobs are composed of two basic systems: (a) The market choice system; (b) the allocation system.
9. For any one country, total outward payments (such as for imports, capital investment and lending abroad) always balance inward payments (for exports and for borrowing and gifts).
10. Man feels a strong need to control those events which have most effect upon his own well-being and societies usually give him some degree of such power through ownership.
11. The more an individual has of material things or power the easier it is for him to gain the next increment of having."

Three types of principles

These principles are regarded by the authors as fundamental to the study of conflict and as independent of value judgement. They are proposed as laws based on logic and experience, being laws of nature, or simple statements of arithmetic, or man-made laws, although it is not always clear which of the principles fall into each of the three categories. Methodologically such identification may be of some importance since a law is, in fact, a model: it implies relationships between several entities which are mutually-exclusive in their definition, so that knowledge of given changes in some will lead us to certain expectations regarding the others. If such a law is a law of nature, then man can predict the expected values of particular variables and the outcome of any actions that he may take within the system governed by that law, he may even be able to determine ways in which he can bring about a desired outcome.

Vol. 4
1976

A man-made law, however, implies acquiescence by all members of society, or at least by all members of a given system, to play by the rules, man-made rules, designed to determine causal relationships and outcomes. Here man can affect the system in two ways, first by his actions (as in the case of laws of nature), and secondly by a changing attitude to the rules (involving attempts to reinterpret, modify or even abrogate them altogether). Thus, prediction of outcomes where man-made laws are concerned, involves an additional element of uncertainty regarding the "rules of the game" and the extent to which they are likely to continue to hold in any given situation. Indeed, the rules themselves may become a function of circumstances and subject to random effects, and this may partly be the reason for man's behaviour being sometimes regarded as totally unpredictable.

A law which is neither a law of nature nor man-made is often easily recognizable as a tautological statement. It is not a law in the sense that it reveals any new relationships between variables, it merely restates definitions proposed for these variables; its main purpose is reinforcement, a useful device in human communications, but the term "law" would be a misnomer, at best we could call it a "tautological law".

This distinction between natural laws, man-made laws and tautological laws underlines the suggestion that the first category is of a more lasting nature, the second is subject to change even in the short term and the third is a matter of semantics. To enunciate laws which are entirely free of judgement and independent of culture, religious creed or political persuasion, one would have to confine one's efforts to natural laws, and the authors of the OPUS document cannot be said to have done only that.

Validity and scope

And the resultant mixture has a symptomatic undulating level of rigour. Principle 2, for example, that when a static cake (total consumption) is divided among several groups, a greater share for one means a smaller share for another, is no more than a corollary of Principle 1 (worth re-stating nonetheless). Principle 5 either implies that man can ascertain his objectives in advance, and that he can (and does) systematically scrutinize alternative courses of action open to him and then behaves rationally in making a choice that will be most advantageous in terms of these objectives—a rather debatable premise—or that whatever outcome he manages to achieve is then defined as his stated objective —in which case we have an example of a tautological law.

Similarly, Principle 10 is somewhat obscure in that it does not properly define ownership: if what is meant is ownership of assets (such as the means of production), then hard evidence is distinctly lacking in support of the contention made; if the term also covers ownership of one's labour, then this is something that individuals have always had and the increasing power of organized labour stems not from change of ownership as from the ability to control union membership; but possibly there is some confusion here between power and responsibility and perhaps the intention of this Principle is first to convey the message that those who exercise power to influence events should develop a sense of responsibility and should be made accountable for the outcome of their actions, since the irresponsible use of power is a recipe for tyranny and exploitation, and secondly to suggest that ownership of assets (as, e.g. in cooperatives) and/or profits enhances responsibility. As for Principle 11 the authors go on to say that society now impedes the divergent flow of individuals to the opposite extremes of wealth and poverty through graded tax systems, death duty and provision of welfare, but they should have added the well-worn phrase "all things being equal". This Principle does not explain why the accepted tendency for the rich to become richer and for the poor to become poorer has not resulted in the total extinction of the middle class prior to, say, 1940 (since the impeding actions of society to which the authors allude have become paramount only during the last four decades or so), and this suggests that all things had indeed not been equal.

But it is not my intention to undertake a meticulous examination of each of the eleven principles and to dissect each at length as to its validity or the rigour

of its presentation. The mixture of disciplines of the unnamed authors may explain the differences in formulation of the principles in terms of style and emphasis, and rather than dwell on the chosen eleven one might wonder about many that could be added to this list, such as:

- man's behaviour is often dictated by his role and not only by his personality
- evaluation of performance can only be done against predetermined targets and yardsticks (or: unless you know where you are going, how will you know when you have arrived?)
- power corrupts; checks and balances are needed to control it
- processes in socio-economic systems are often irreversible, and when they are reversible there is a hysteresis effect (or: things are not what they used to be)
- there is always a conflict between the short term and the long term (this is perhaps a more general statement than Principle 3)
- most industrial situations are not zero-sum games and strategies of envy or maximization of self interest are not always best in terms of outcomes

Rational man

One could go on, and the authors of the OPUS document would undoubtedly be delighted to examine new propositions and add to their magnificent eleven. But it seems that a much more fundamental question needs to be asked: What is the precise purpose of these principles? If it is to enlarge our knowledge and understanding of society, then this is laudable enough. But the OPUS pamphlet is much more explicit in its objective, and apart from urging all responsible people in society to communicate these proposed principles to the populace at large, it ends up with the statement: "A better understanding will reduce conflict and release additional resources of energy to contribute to a more constructive and happier society."

Vol. 4
1976

This is a typical positivist–rationalist view, commonly held by students of the industrial scene; it is the school of thought that believes that problems facing society can be formulated and analysed in a systematic way, that when a number of courses of action are available it is desirable to examine them in terms of their relative merits and outcomes and that it is therefore possible to identify the best alternatives as judged by pre-defined criteria, that once an optimal solution to a given problem is determined it should be embraced and implemented—in short this school of thought believes in rational man.

What we need to ask ourselves is first, whether rational man is a realistic proposition, and secondly whether—assuming rational man is adopted as a working hypothesis—the result would be an elimination or at least a reduction of conflict. As for the first, it involves not only the assumption that rationality is universal and impersonal, that it is capable of a dispassionate determination in each case, that it is culture-free and that what is seen as rational by one individual will be readily (or after discussion) accepted as rational by another,

but also that once an individual recognizes what is a rational thing to do he will indeed do it. We know, alas, that neither accords with everyday experience. Arguments do persist about what solution logic should dictate, and no analysis of socio-economic and managerial problems is entirely value-free. Whether this state of affairs stems entirely from lack of agreement on goals (where it is not difficult to see that value judgement is bound to play an important part), or whether the choice of analytical tools also contributes to a potential discord, is a matter for debate; but the fact remains that we often encounter disparities in views, both concerning how problems should be defined and how they should be solved. And with regard to human behaviour, deviations from what appears to be rational choice are observed often enough, either because of inadequacy or ambiguity in goal definition, or for reasons of personality and other forces which may even be unconnected with the goals in question.

The second question alluded to earlier is whether rational behaviour is bound to reduce conflict. If we adopt a rationalistic viewpoint we have to accept that when a problem has a single optimal solution and if the interested parties come to realize and accept its logic, any conflict between them will automatically be dissipated. But it is similarly inevitable that when there is no single solution, that when one party can inflict incalculable damage on another in the pursuit of its own interests, it will indeed proceed to do so, and under those circumstances (which often occur in non-zero-sum games) rationality in the sense understood here will only exacerbate conflict rather than abate it.

Conflict theories

It is indeed curious how diverse are the prevailing views on the question of conflict and its resolution, as suggested by the following spectrum:

1. the one happy family theory
2. the paternalistic theory
3. the consensus theory
4. the conflict resolution theory
5. the continuing conflict theory

The first is held by those who are perturbed by the whole notion of conflict and who seek to avoid it either by denying that it exists or by trying to dismiss its implications. They argue that there is much more that unites employers and employees than divides them, that in the face of the overwhelming importance of their common goals sectional interests shrink into insignificance, and that the "working together" ethic is an essential element in the maxim that prevention is better than cure. And as long as prevention works, this theory is commendable enough, but it is impotent when conflict does arise because it does not explicitly address itself to the problem of distribution of profit and power in the industrial enterprise.

The paternalistic theory takes the view that an industrial conflict generally involves two very unequal contestants, so unequal that one can be completely

crushed by the other. The dominant contestant must, therefore, be aware of his power and show compassion by voluntarily allowing the weaker side to obtain greater benefits than those that he would have to be content with if judged solely by the power balance in the struggle. In years gone by it was the employers who had to exercise such a paternalistic role (albeit not always wisely); in modern times the power balance has shifted quite considerably, but those who hold to the paternalistic theory would argue that its message remains valid irrespective of which side is dominant in a particular struggle.

The consensus theory looks for a resolution of conflict in a way that is generally thought to be fair. It maintains that justice need not only be done but must also be seen to be done, that it is neither acceptable for the strong to impose their will on the weak nor is it sufficient for the strong to be magnanimous in a paternalistic fashion. Furthermore, a consensus of what is fair is not a matter to be confined to the parties involved in the conflict; since it is rare in modern industrial society for one sector to be totally immune from the repercussions of what happens in another, it is essential for outsiders to the conflict (outsiders with varying degrees of objectivity) to accept that justice has been done. It is not, of course, easy to determine in each case what the consensus is—the press and broadcasting media are both a help and a hindrance in this respect—but, as believers in the consensus theory will argue, this is a technical problem, which does not in itself invalidate the need for consensus; without it, resentment and enmity may ensue, sowing the seeds for fresh and even more bitter struggles in the future.

Vol. 4
1976

A consensus solution however, is not necessarily a rational solution; since a consensus inevitably involves taking the temperature of a political climate, particularly when major issues are involved, the recommended solution may well be influenced by strong emotions and transient considerations. Rational man will argue that no solution can have a lasting value unless the problem has been looked at dispassionately, that if reason is not satisfied, objections to the solution are bound to persist. It is far more profitable, therefore, to resort to logic and solve a given problem once and for all, in the same way that we solve mathematical problems under given premises. This conflict resolution theory has its attractions, but it suffers from two shortcomings: first, industrial problems (unlike mathematical problems), rarely have generally recognized premises and agreement needs to be reached about these premises before an analytical examination can be undertaken (and, alas, people's preferred premises may be affected by their expectations of the resultant solutions), and secondly for many non-zero-sum games there are no agreed analytical solutions, and questions of judgement and personal preferences regarding methods of co-operation between the players may thus remain.

Finally, the continuing conflict theory is not suggested as a substitute for the other theories, but as a supplement. It simply points out that a conflict is never finally resolved, that as soon as a solution is adopted for a given situation new

problems arise; and either the conflict re-emerges under a new guise or new conflicts are generated. It is therefore idle to pretend that any given conflict is an isolated static affair; it is but an element in an evolutionary process and hence it follows that whatever method is used to resolve a particular conflict, any bargaining that is employed, any consensus that is sought, should take account of the dynamic nature of conflict development and the changing forces that shape it.

Having said all this I must again concede that, while I continue to have reservations about the particular set of the OPUS principles, a case can obviously be made for attempts to improve our knowledge and understanding in this area. There is no guarantee that such knowledge will always be used to the good, indeed it would be surprising if it were not deliberately misused by the greedy and the unscrupulous, in the same way that all knowledge has been misused in one way or another during the history of mankind. There is no guarantee, therefore, that society will be any the better for knowing how it works, and there is certainly no quarantee that industrial conflict can thereby be reduced. And thus, in the application of whatever knowledge we manage to gain, it is well to beware of the pitfalls inherent in any of the theories of conflict resolution and to counsel caution. Attitudes of rationality, tolerance, moderation and compassion are not always compatible, yet they all need to be exercised in the management of industrial affairs; alas, not one of the theories can claim to encourage all these attitudes to such a degree that it could be totally embraced on its own to the exclusion of all others.

SAMUEL EILON
Chief Editor

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

FEEDBACK

Methodological Optimality in Operations Research

PROFESSOR EILON'S editorial, *How Scientific is OR?*, [2] addresses the similarities and differences between the hypothetico-deductive version of "the" scientific method and the OR process. In doing so, he provides ". . . food for thought, both to those who insist that OR is in every respect a scientific activity, and to others who may prefer to regard it as an art".

Without engaging in the age-old debate as to the degree of "art" that is inherent even in "traditional" science, I should like to propose that OR is, by virtue of its heritage as an approach to solving *operational* ("real world") problems, intrinsically and necessarily a complex mixture of art and science. Moreover, I would like to propose an operational criterion for methodological optimality in OR which incorporates both the "artistic" and the "scientific" aspects, at least as they regard that most important area of concern to operations researchers—*implementation*.

Professor Eilon's "OR Process Model" recognizes the great significance of implementation, but nonetheless treats implementation considerations in OR as elements that are necessarily *preceded by* modelling activity, and probably, also preceded by data collection and analysis. Such a view is, in my judgement, one of the factors contributing to the many documented and undocumented failures to implement "optimal" OR solutions that have been so widely discussed in recent years.

If this is so, there is a clear need for a broader concept of "optimality" than that which operations researchers—and scientists in general—have traditionally used. This broader concept of optimality is *methodological optimality*—that optimality criterion which the operations researcher applies to his selection of the "best" model, that "most appropriate" data collection procedure, and the "proper" analytic technique to be used in the problem-solving process. This methodological concept of optimality is one which goes beyond simple "technical optimality" to *incorporate implementation considerations into every step of the problem-solving process.* The underlying premise of this broader concept of optimality is simple: that it is far better to develop a solution which improves a system, and which is implemented, than to develop one which (theoretically) optimizes the system's operation, but which is not implemented.

This view of methodological optimality is clearly based on a view of OR as primarily an "applied" science since it devalues theoretically-optimal solutions which are not implemented. Indeed, it even goes so far as to devalue solutions which were not developed and designed specifically to have a high likelihood of being implemented, even if those solutions are, in fact, implemented in a particular situation.

This view reflects my personal values, which are clearly not shared by all of my OR colleagues. However, as I review the body of material that has been given the title of "OR" over the past thirty years, it is difficult for me to discern *anything* that distinguishes OR from such traditional disciplines as economics and "applied" mathematics other than its real-world "implementation" orientation. Thus, on the simple pragmatic basis that if nothing other than this implementation focus serves to differentiate OR from older, better-established and more highly regarded disciplines, either OR must address itself directly to this "competitive advantage" or it should, and will, disappear.

The basis for a broader methodological concept of optimality

The idea of a broader concept of optimality

which incorporates implementation considerations into every stage of the problem-solving process is consistent with much of the implementation-oriented thinking which has gone on in recent years, as exemplified by Churchman and Schainblatt's "principles" regarding the needs of managers for information which suits their psychology and for methods of generating evidence which is geared to their problems.[1]. It would appear to be generally agreed that except in very special circumstances, if an OR analyst and his client-manager do not have such "mutual understanding" and if the information (solutions) provided to managers does not suit their psychology and needs, a good manager is unlikely to be willing to act on the basis of a recommendation made as a result of an OR study.

Indeed, in a closely-related field—that of management information systems (MIS)—the analogous implementation problem has been viewed as such a function of lack of understanding between managers and analyst that a high-level committee of the Association for Computing Machinery has recommended the creation of a new MIS job position. The primary function of the new position is to serve as an intermediary between manager and analyst, so that their lack of understanding of each other's roles and methodologies will not serve to prevent the successful implementation of information systems.

Thus, the old controversy about whether better implementation is to be achieved by developing better, more analytic, managers who are able to understand the sophisticated models and computer systems of the analyst, or by giving analysts training in management to ensure that they understand the complexities of the real-world manager's job, is apparently to be resolved in the MIS field by giving up on both possibilities and interposing an individual with an entirely new variety of skills to operate between manager and analyst.

Another much-discussed way of developing a better manager-analyst relationship and of ensuring that the informational output of an OR study is "needs" oriented is to incorporate the manager into the problem-solving process—the behavioralist's panacea of "participation".

Neither the idea of creating a whole new field of study nor the presumed, but largely untested, beneficial results of participation would appear to resolve the basic dilemmas of implementation in an effective and satisfactory way. The interposing of another individual between the manager and analyst may simply hamper communication and will certainly increase the already-high cost of analysis.

I do not decry the idea of participation as a beneficial approach; indeed I believe it to be important. Whether participation is achieved within the framework of Schultz and Slevin's [5] *Behavioral Model Building*, King and Cleland's [4] "manager-analyst teamwork" or otherwise in less structured fashion, participation is a necessary, but not sufficient, condition for successful implementation of the "average" OR project.

Before going on to further operationalize the concept of methodological optimality that will permit participation to be really effective, let me discourse on this use of the term "average". This entire discussion is oriented toward the "typical" or "average" operations researcher and OR project in the sense that I recognize the special abilities of a very few OR practitioners to succeed using techniques that would be failure-prone for most of us. These few analysts have the unique ability to instill in the manager such a degree of *personal* confidence in the analyst that "understanding" on the part of the manager is overshadowed in importance. The manager accepts analytic results on the basis of his faith in the analyst's abilities, and so long as actual results are commensurate, his faith is reinforced. However, since the vast majority of operations researchers of my acquaintance do not have, and would not claim to have, such abilities, it is necessary to develop an OR problem-solving process which can serve all.

The methodological optimality concept which can do this is one which does not merely rely on vaguely defined "participative" and "understanding" as the path to successful implementation. Rather, it is one which builds on mutual understanding and participation to encompass an explicit agreement on the part of both manager and analyst that *every phase of the process is to be judged both on the basis of technical optimality and on the basis of its "contribution to implementation"*.

This criterion extends to the selection of a model as well as to data collection and analysis. Thus, using this concept, the model which is jointly chosen to be best by the manager and analyst may not be best in the

Vol. 4
1976

technical sense. It may not, in the case of a regression model, have the highest R^2, or, it may not, as suggested by Eilon's diagram of the OR process, be that model which is best at prediction. Rather it will be a model which is good enough in these technical senses *and* one which is likely to be implemented.

The implications of operationalizing a broader approach to methodological optimality

If this description of a broader approach to methodological optimality is viewed as merely a pontification, most operations researchers will probably agree, "in principle" However, if it is viewed as a practical approach to modelling and problem-solving, much concern and debate will undoubtedly arise concerning its implications.

First, there is the obvious concern that "likelihood of implementation" will be operationally translated into the degree to which the model and process confirm previously-held beliefs of managers. This would mean that only those model elements to solutions which confirmed the manager's prior judgement about the world would be designated to be implementable. Clearly, carried to the extreme, such an approach would restrict organizational change and innovation and the organization would stagnate. Just as clearly, it is one of the responsibilities of the analyst to see that this does not occur.

One way that the analyst can do this is to be insistent that the rationale for *not* using the theoretically best model, data collection procedure, or data analysis technique be made explicit and be evaluated. If he is so insistent to the same degree that most good managers are insistent that a proposed model be explained and defended by the analyst, the process will be a valuable one in enhancing the understanding of both manager and analyst and will prevent degeneracy toward previously-held notions.

Another implication of this approach has to do with the trade-off between descriptive and normative models. OR consultants often construct descriptive models of organizational procedures and decision-making as a basis for developing their own understanding. They view these models as constituting merely a step along the way toward the development of normative models. However, with this broader concept of methodological optimality, the "how-we-are-used-to-doing-it" element plays a more vital role. Indeed, using such a concept, one may find it useful to develop a formalized procedure for explicating and resolving conflicts between descriptive and normative models. For instance, King and Cleland [4] use a "consensus" model, which is directly based on explicit normative and descriptive models, as a means of balancing the theoretically best model with the current "state of the art" in the organization. If the consensus model is so developed in a participative fashion, it should approximate the broader variety of methodological optimality and thus permit practical improvement in the organization.

The implications of this approach in terms of the skills required of the OR analyst are also significant. Using this concept of optimality, he cannot play the role of the "pure" detached scientist. He must subtly use his persuasive powers to urge the manager and organization *in the direction of* using the theoretically best model, or some reasonably close approximation to it, of collecting the proper data and analyzing it in the appropriate way. He is no longer able to simply prescribe that model which he deems to be best or that data collection and analysis procedure which he deems to be proper; rather he is in the position of a guide who will probably be forced to accept something other than that which he advocates.

This is a significant broadening of responsibility for the OR analyst who may prefer to operate in the relatively sterile environment in which he makes his recommendations, has them accepted or rejected, and moves on to new problems. This proposed *modus operandi* places responsibility for the outcome of his analysis—its implementation—on the analyst, just as managers are responsible for the outcome of their decisions. Such is the basis for the development of true "mutual understanding".

The pure scientists which lurks beneath the breast of many operations researchers may find such responsibility and such a role to be difficult to accept.

REFERENCES

1. CHURCHMAN CW and SCHAINBLATT AH (1965) The researcher and the manager: a dialectic of implementation. *Mgmt Sci.* **11** (4), B69–B87.
2. EILON S (1975) How scientific is OR? *Omega* **3** (1), 1–8.

3. KING WR and CLELAND DI (1971) Manager-analyst teamwork in management information systems. *Bus. Horizons*, April, 59–68.
4. KING WR and CLELAND DI (1975) The design of management information systems: an information analysis approach. *Mgmt Sci.*, to appear.
5. SCHULTZ RL and SLEVIN DP (1972) *Behavioral Model Building*. University of Pittsburgh Working Paper No. 72–1, Pittsburgh, Pennsylvania.

WILLIAM R KING
(9 July 1975)

Graduate School of Business
University of Pittsburgh
2125 *Cathedral of Learning*
Pittsburgh
Pennsylvania 15260 *U.S.A.*

Management Careers for Management Scientists—Comments

I MUST take issue with the implied premise that management as such is something that can be taught. I feel that in our present society we have too great a tendency to regard management as a profession rather than a philosophy, with too much concentration on the administrative side of management instead of on aspects of risk-taking, innovation and initiation. The management functions can be considered as follows: (a) planning, (b) organising and delegating, (c) co-ordinating, (d) motivating, (e) controlling, (f) initiating and innovating. The management apprentice scheme concentrates on items (a) to (e) i.e. the administrative skills, and in many respects this formula is repeated in general management education, whether it be through academic institutions or management training centres.

By all means managers must be capable administrators, but to be successful they must also be self-starters. This is my main point. Successful managers must be leaders, innovators, initiaters and the source of management talent is not important provided the individuals have these qualities and can be taught the administrative skill. The important point is that talent, however unorthodox, is recognised, encouraged and allowed to flourish.

In many respects the Management Science environment is a better training ground than the apprentice scheme or management institution because problem-solving in many cases requires an innovative approach, thus allowing a potential management candidate to be recognised. It would however, be wrong to assume that other corporate areas are barren of managerial potential.

The "barriers" in the way of advancement cannot easily be overcome. The management scientist has it in his own control to recognise the importance of less intellectually stimulating work and to accept loss of status in a sideways move as a form of investment in experience. Firms in their turn should be prepared to give responsibility and status early, and to break down barriers of tradition. The "tradition barrier" is perhaps the most difficult to break down, since firms of this traditional nature tend to be administered rather than managed, and have stifled potential early on.

JG HOWARTH

Brooke Bond Oxo Ltd
Leon House
High Street
Croydon CR9 1JQ

PROFESSOR EILON'S Editorial (*Omega*, Vol. 3, No. 3) draws attention to the problems of career development of MS (management science) staff, and in particular to the problems of movement of analysts out of MS groups. Undoubtedly most organisations with an established MS group have to some extent experienced the problems you identify. I am not so sure, however, that the situation is as difficult today as it was a few years ago.

Eilon singled out three principal barriers to mobility from an MS group. Consider first the salary barrier; whilst it is true that ambitious and energetic management scientists can make rapid progress in salary and status, the same can now be said for many

Vol. 4
1976

other groups like planners, economists, and currently, young accountants. Consequently, other functions can offer positions to management scientists which do not necessarily involve a sideways or downward move to break away from the MS group.

The second barrier, that of tradition, does of course exist in organisations but here again this is less forbidding than it used to be. In recent years my own organisation has made significant progress towards career development based on performance and potential, with perhaps less importance being attached to rigidly prescribed experience.

Lastly, the intellectual arrogance barrier is perforce being removed, as management scientists have learned that the ability to communicate and deal with people at all levels is as important as technical ability for the effective use of their skills. I believe that in recruiting MS staff, managers are placing as much emphasis on these personal qualities as intellectual capacity.

The above should not be taken as a complacent view of the problems, but I feel the position is not as intractable as the Editorial would seem to suggest.

BP HUGHES
Marketing Research and Management Sciences
Shell-Mex and B.P. Ltd
P O Box No. 148
Shell-Mex House, Strand
London WC2R 0DX

I HAVE two main reservations concerning the line of argument Professor Eilon uses in his Editorial on "Management Careers for Management Scientists" (*Omega*, Vol. 3, No. 3).

Firstly, I would question the suggestion that management education at universities "... is concerned with ... management as it should be". Rather would I see it as the educationist's attempt to provide some of the ingredients currently thought to be useful to a manager. Until we understand more about what constitutes management and, secondly, how best to teach it, I doubt if we can claim more than that. There is no "one best way". Furthermore, I very much doubt if employing organisations would uniformly prefer to obtain all their management recruits from amongst the students of such departments, even if they were available in sufficient numbers. Though some of this reluctance may be due to a preference "to recruit in one's own image", I fancy, that some organisations have preferred to develop personnel policies, which rely upon some mix of the various methods of entering management, to give them the kind of variety and interplay between individuals, that they think is beneficial. This is more than just an attempt to hedge their bets, since, in addition to some overall assessment, subsequent management development programmes to some extent allow comparative assessments of the value of the various methods of entry. In particular, therefore, it cannot be taken for granted that there is no better training for management than a two to four year period in a management science group, as you suggest.

Secondly, I should like to add at least one category to Eilon's three categories of management scientists, to cover those who do not necessarily want to stay in the MS group, but are not considered to have the qualities which would allow them to transfer into management. There is surely no simple link between the suitability of a person to tackle problems off-line, and his suitability as a manager. In addition, I would imagine that, for example, the more craft-based and professionally technologically-based of industries in particular would tend to rely more upon other forms of entry into management. For example the former industry might prefer some form of "learning through experience" as a basis for their management training, rather than upon recruitment via their MS groups (as opposed, perhaps, to the occasional secondment of a suitable management trainee to the MS group); whilst the latter might restrict recruitment into management primarily to its particular type of professional technologist.

J BANBURY
Industrial Sociology Unit
Imperial College
Prince's Gardens
London SW7 1NA

I FIND little to object to in Eilon's general analysis of the three routes to the early stages of a management career, the trainee scheme, the general M.B.A. type university or polytechnic education course, or the more specialist

route through one of the management related subjects or functions, exemplified by management science in Eilon's discussion. Possibly his description has a slight aura of yesterday in its implication that organisations take on graduates and others into safe careers, where if one behaves reasonably competently and decently one has a career for life made up of a carefully planned series of jobs. All too many organisations have found themselves forced by circumstances into reneging on this particular psychological contract, experiencing the agonies of redundancy programmes, and beating a rather disorderly retreat from "you fit in with us and we'll look after you" to "it's *your* career, and it's up to you to make sure that you keep equipped with skills that have a market value". This creates a new dilemma for those responsible for managing recruitment and training of managerial staff: how to encourage loyalty without instilling complacency and taking on moral commitments on behalf of the organisation which it cannot be certain to be able to fulfill. All this makes the career system rather more organic, and nearer to the law of the jungle where the fittest survive. To the extent to which this is the case, management scientists should have nothing to fear, if this kind of work is the excellent training for management that you claim it to be. With their broad experience, analytical skills, independence and political awareness they should be one of the predominating species. I hope that the Editor, as a scientist, is pleased that circumstances are providing a natural experiment to test his assertions!

In case you are beginning to suspect me of mild sarcasm at this point, let me assure you that I share the Editor's hope. Although we must face the fact that we are currently going through a period of general loss of faith in science as a socially useful process, it is still apparent to me from my direct observation of management practice that there is enormous scope for improving efficiency and effectiveness through the application of the most simple, rational and systematic procedures. Surely the one-eyed man should triumph in the country of the blind. . . .

However we must recognise that Eilon is pleading a special case for management science, and that there are other subjects/approaches/functions which serve as a preparation for, contribution to, and ways into, management. They could, and I am sure do, argue in just the same way, that they are excellent, if not the best, preparation for management just as you do. I am sure they will not thank me for amateur attempts at their arguments, but I can imagine the following:

> The Marketer: "Marketing is a uniquely valuable preparation for general and senior management for it is the function committed to keeping the organisation in touch with the outside world, and it is on senior managements' ability to adapt their organisations to changing circumstances that survival and prosperity depend."
>
> The Accountant: "Cash is the very lifeblood of any organisation, and finance the natural language of business. Experience and skill in finance and accounting provide an ideal background for senior management, since they offer the only way of keeping in touch with progress in various parts of an organisation, without losing sight of its overall performance."

It is a widely observed phenomenon that everyone believes there is a special superiority in the groups with which they identify themselves. It always amuses me that in our field every management related discipline or subject has a model using its own concepts and terms that includes all the other disciplines and subjects, and we play an endless game of "my model includes your model", "no my model includes yours" (which is not to deny that this may be a very fruitful way of developing ideas)!

Recognising this phenomenon does not stop one being part of it, and, remaining faithful to my "behavioural" origins I must point out that this phenomenon of group identification involves acquiring the values, norms and beliefs of one's reference group or groups. This is the process Eilon refers to when you pin-point the aspect of organisations' management trainee induction processes which involves "an introduction to the firm, its organisation and *ethos*, *often involving subtle* (*and, sometimes, not so subtle*) *indoctrination* . . .". Indoctrination is a negatively valued word in our culture, and the Editor's statement carries the implication, either intended or not, that there is something undesirable or disfunctional about this process. I argue that this process is inevitable and necessary if any group, be it organisation or professional grouping like accountants, organisation theorists or management scien-

Vol. 4
1976

tists, are to be capable of any cooperative work. I am sure that the process of becoming a management scientist must involve such a process of socialisation, and suspect that the Editor's approving quotation from Keynes concerning rejection of liability to obey general rules, reserving the right to judge individual cases for oneself, gives a flavour of the predominating values involved here.

If Eilon can accept that there is a degree of truth in this analysis of the situation, then we have a new perspective on what I take to be the central issue to which he is addressing himself: how can some of those who start their careers as management scientists make a role change in the direction of becoming a manager, particularly a senior, general, rather than functional, manager, where Eilon argues convincingly he would have the skills and experience to contribute particularly well to organisational effectiveness. It seems to me, on the argument I have devel: oped above, that a major barrier to this role change is the natural difficulty the person will have in adopting the new values that will be necessary to allow his acceptance as a "manager". This analysis does not offer an easy solution to the problem, perhaps only the suggestion that the transition might be easier if the people concerned know that this is what is going on, and that arguments about whether or not a person has the necessary skills and experience for management, are often only in fact about this on the surface, and the real issue is whether he has acceptable attitudes and values.

However it seems to me, that we, (and in using the term "we" I invite the Editor to identify with me in a common group of people devoted to improving management through teaching, research and consulting) at least partially create this problem, through our own divisiveness in working hard to make our approaches to helping management stand out from others, so that we are organisation theorists, corporate strategists, marketeers or management scientists first, and management educators/researchers second. Perhaps we would do ourselves, and more important our students and organisational clients, a favour, if we spent more time looking for similarities and common ground, and less looking for differences. If the Editor's response to this is to say "management science has the map of the common ground and the method to navigate from it" my response is "just as my behavioural analysis predicted", and there we are back in our boxes. But if the Editor points out that all I have really been doing is to try to talk you over into my box, I might have to plead guilty, and this perhaps identifies the nature of our dilemma.

JOHN G BURGOYNE
Management Teacher Development Unit
University of Lancaster
Lancaster LA1 4YR

PROFESSOR EILON has reviewed the problem of assimilating, training and developing management scientists within industry in his Editorial of (*Omega*, Vol. 3, No. 3). He has rightly pointed out that industry has been largely unsuccessful in making best use of graduates. This failure to make best use of their most valuable resource, professionally trained management potential, has been particularly acute with engineering and scientific graduates. However, similar criticism can be levelled at industry's attempt to develop and make best use of social science graduates and indeed, those most specifically educated in preparation for industrial management, the business graduate.

To my knowledge, the classic training formula referred to by Professor Eilon, has never been entirely successful as measured in terms of staff retention ratios and job satisfaction for the individuals subjected to them. A review of the traditional four step assimilation and training formula (management skills courses, corporate indoctrination, the "milk round", and project assignments) is long overdue and some progressive companies (probably those with senior management who have been subjected to such training in their early careers) have been developing more enlightened methods aimed at assimilating, training and developing graduates so that they may successfully progress through a management career, whether it be in line or staff, or both.

I would like to outline one method of approach which has been used with a great deal of success both in the United Kingdom and abroad. The key to the success of this approach is in the effective recruiting of suitable graduates for particular line functions.

Pre-recruiting: Before actual recruiting activity begins, all line and staff managers are requested to assess requirements of graduates for the coming year. These positions are generally of junior management or supervisory status. Departmental managers are requested to write job specifications and plan, in detail, the role of the intended recruit.

Once the total requirement of graduates has been determined for the corporation, a carefully designed advertisement is placed in appropriate media in order to reach the target groups.

Mutual selection: Selected graduates are invited to participate in a series of interviews in an atmosphere which provides the opportunity for corporate managers to be questioned about jobs and various aspects of the company. These interviews are a critical part of the selection process as they take place at a number of unstructured exposures with management from various parts of the corporation, both line and staff. Potential recruits have these meetings mostly with groups, informally at luncheons and also at group discussions. Here graduates feel much more free to ask probing questions about the company and the jobs being offered. As the selection programme takes place over about five days, both parties have a much better than average chance of assessing the suitability of the proposed partner. Line managers then follow up with specific job offers to particular graduates thought suitable for the particular line functions discussed during the interview week. If the approached graduates were impressed with the specific function, and what they have learnt about the company in talks to the various management, they take up the offer knowing exactly what the line position responsibilities will be.

Training and development: Once recruited, the graduate undertakes on-the-job training. Training is largely devoted to developing the skills of his existing supervisory junior management role. However, training is developed in perspective of the position and line department in relation to the corporation.

There are, in fact, three ways graduates may progress through the organisation. Firstly, there are promotional opportunities within the line department they were recruited into. Here they are in an excellent position to be primary candidates as they have had both graduate training and the chance to develop as a line manager on the job. Experience in this method of approach has shown that working oneself up from junior management positions, which is often quite rapid, almost invariably overcomes the Tradition Barrier that Professor Eilon refers to. The demands of the graduate's first junior line position also dilute any feeling of superiority the recent graduate may enter the organisation with (Professor Eilon's arrogance barrier).

Secondly, there is the opportunity to progress to a more senior management position in another line department. In practice graduates with line experience in a specific area, are highly sought after in other line areas. Here, the knowledge built upon the previous department is often important. The senior manager in the new department will recognise that the graduate will learn the function quickly and be able to build on his earlier management experience and company knowledge.

The third avenue for progression is into a staff role or project team. In practice, line developed graduates are highly sought after for such positions. Not only can they bring their scientific training to bear on the project or staff task, but they have also by this time developed a working experience of an important part of the company's day to day operation. Their recommendations are likely to be more practically oriented than a new graduate or narrowly exposed staff/project man. In addition to this, recommendations made by them will have a much higher credibility when action decisions are being made.

Companies which have tried this new approach of assimilating graduates through initial line exposure have had rewarding experiences as far as on-the-job performance and retention rates are concerned. One company has had a 90% retention of graduates over a two year period. Graduates developed through line experience have provided an excellent source of management material at various levels, both for line and staff departments.

Graduates assimilated in this way have been integrated into the company much more satisfactorily under this new approach. Many of the frustrations caused by the traditional "separate" treatment approach have been eliminated or minimised.

The initial experience of the graduate after joining a company is all important in terms of his own assimilation and development.

Vol. 4
1976

Japanese companies call this period *gasshuku renshu* (living-together practice). In Japan, it is all important that assimilation is as effective as possible, as the newly recruited graduate will be expected to remain in loyal service to his company until he reaches his fifty-fifth birthday when he receives a lump sum payment, which will serve both as a pension and a reward for his long and faithful service. Although western enterprises are not likely to develop the graduate–company partnership to this extent, a much improved induction approach will return handsome dividends to both employer and employee alike.

BRUCE E PERROTT

5 *Kingswood Avenue*
Hampton
Middlesex

l. 4
976

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Energy Analysis: A Review of Methods and Applications

PETER CHAPMAN

(*Received June 1975; in revised form September 1975*)

The paper describes the aims and applications of energy analysis, emphasising its role as a complement to conventional financial analysis. It is shown to provide useful insights in applications such as forecasting price rises due to a fuel price rise, calculating the net costs of fuel obtained from unconventional sources, forecasting energy demand and evaluating energy conservation schemes. Although the field is still young, and most activity is in compiling reliable data bases, its systems approach and methods promise to make it a useful tool for evaluating patterns of resource uses.

1. INTRODUCTION

ENERGY analysis is a systematic way of tracing the flows of energy through an industrial system so as to apportion a fraction of the primary energy input to the system to each of the goods and services which are outputs of the system. The result of an energy analysis is a set of 'gross energy requirements' (ger) for all the commodities produced. When applied to a national economic system then the quantities of all the commodities produced, each multiplied by its respective gross energy requirement, will add up to the primary fuel input to the system. Thus, an analysis which concludes that the ger of a loaf of bread is 20 MJ[1] implies that the production of a loaf of bread required the removal of 20 MJ of primary energy from some energy resource stock. Some would be used in extracting the fuel and converting it to a form suitable for use, some would be used in transporting commodities between various stages of production, some would be used in producing materials and equipment used in the processes and some would be used as a direct process input, such as the fuel for ovens.

Energy studies were given a large impetus by the oil price rise in 1973 so that although the methods of energy analysis were used prior to 1973 most studies have been carried out in the last 2 years. The very rapid development of the subject has given rise to misunderstandings as to its aims and range of application. The purpose of this paper is to clarify the aims and applications of energy analysis. To accomplish this it is necessary to discuss the methods of analysis and the conventions used (Section 2); to relate energy analysis to financial

[1] Note that 1 MJ = 0·28 kWh = $9{\cdot}5 \times 10^{-3}$ therms = 950 Btu.

analysis (Section 3); and to discuss some of the applications of energy analysis (Section 4). This will make it clear that energy analysis is not an *alternative* to economic analysis but a way of refining certain parts of economic analyses. Energy analysis can complement economic analyses since it can be directly related to the state of technology and its methodology involves examining the whole industrial system, not a single firm. These two aspects of energy analysis are illustrated in the examples given and used to give some information about future costs.

2. METHODS

There are three basic methods available for evaluating the gross-energy-requirements of goods and services, namely statistical analysis; input–output analysis and process analysis. These have been described in detail elsewhere [10, 11, 19] and their relative merits discussed. Briefly the problems and advantages of these methods are

Statistical analysis: Uses published statistics as data source. Able to compare technologies where suitably disaggregated statistics exist. Not usually possible to find comprehensive statistics on all inputs to processes except for census data. Published statistics may contain significant errors, especially due to double counting [2].

Input–output analysis: Based on published economic input–output tables. Degree of aggregation fixed by the table, cannot compare different processes for same commodity. Errors due to preferential prices and statistical adjustments needed to balance table. Advantage—that it takes into account all indirect effects. This method gives the ratio of the total energy input to the financial value of the output, referred to as the energy intensity, for an industry or commodity group [24, 49, 50].

Process analysis: Based on detailed evaluation of process inputs often in co-operation with producer. The only technique for comparing processes and technologies in detail. Problems in deciding where to draw system boundary [10, 19]. Needs data from one of above methods to determine ger of minor inputs (otherwise one process analysis leads to analysis of all production).

In August 1974 an international workshop considered the methods and conventions of energy analysis and agreed a basic set of ground rules which are common to all the above methods [19, 38, 46]. However, these conventions are necessarily incomplete and any particular study must add its own conventions, based on the aims of the particular investigation. The most important factors which have to be chosen are those relating to

(a) the definition and energy content of primary fuel inputs: should solar energy be counted? should uranium be counted on its theoretical energy content?

(b) the definition of the system analysed: in particular this requires specification of the ger of imports.

Vol. 4
1976

(c) the partitioning of energy used in co-production processes or common services: this is particularly important if the problem posed involves changing the quantity of output for just one of the co-products.

The fact that different investigations, seeking answers to different questions, may use different conventions in the above respects means that care is necessary in interpreting the results of energy analyses. The conventions used in an analysis are *not* arbitrary but must be chosen to suit the definition of the problem under investigation [10]. For example in evaluating the ger of a loaf of bread some convention is needed to establish the ger of imported wheat. If the study has the aim of evaluating the fraction of U.K. primary fuel used in the production of bread then the only energy requirement is for transport of the imported wheat from dock to bakery. However, if the study were evaluating the total energy sequestered from the world stock of energy resources for the production of bread then the imported wheat should be given a ger which reflected the energy expended in its production and transport from the exporting country. Clearly these studies will result in different numbers for the ger of a loaf of bread. There is no sense in which one number is more 'correct' than the other, but the subsequent application of either result would have to be consistent with the convention used in its evaluation.

A similar problem of interpretation has arisen as a consequence of energy analyses of nuclear power stations. At least one author has seen the confusion and misinterpretation of results as sufficient grounds for questioning the validity of energy analysis [28]. However, it is clear that an investigation which asks "how much fossil fuel energy is required to build a nuclear reactor?" will produce a different answer to one which asks "how much fossil fuel will we save by substituting a nuclear power station for a coal-fired station" since the latter question presumes a substitution not considered in the former. Provided these differences are taken into account then all the analyses of nuclear power are consistent.

Recently Hill and Walford [25, 28] have attempted to divide energy analyses according to the aims of the investigation. This has led them to suggest that there are six *methods* of energy analysis. This classification, which confuses methods and aims, is not helpful since in practice most analyses subscribe to more than one aim and use most, if not all, of the methods distinguished by Hill. A significant feature of most energy analyses carried out to date is that they are mostly concerned with evaluating the average ger of a good or service. However, this may be inappropriate when the energy analysis results are to be used in comparing alternative policies which involve change. For example the average ger of passenger transport by road and rail incorporate the present load-factors of both systems and suitably amortised fractions of the ger of capital items such as roads and rolling stock. A small shift of passengers from one transport mode to another might not incur any change in capital and may

simply change the load-factors of both systems. Under these conditions the change in energy consumption would be more accurately calculated using marginal energy requirements, not average energy requirements. If policies were contemplated which involved a massive shift of passengers from one mode of transport to the other, then it is likely that the most significant factor in energy use would be the energy required for the immediate capital investment [11, 12]. Thus for policy applications the present practice of evaluating average ger is not appropriate. This is an area where the methods of analysis and presentation of results could be significantly improved.

3. RELATION TO FINANCIAL ANALYSIS

A conventional financial analysis considers costs in terms of factor payments. In addition to payments to labour and capital there will also be payments for items purchased from other suppliers. A new aluminium smelter would include payments for electricity, bauxite, carbon anodes, etc. This procedure takes no special note of how the suppliers of these items arrive at their costs (or prices) so it is impossible to determine how sensitive the total cost estimate is to commodity prices. For example, increasing the price of oil or the average wage rate would affect all the items listed above, but each to a different degree.

Vol. 4
1976

In the conventional analysis the system studied is the firm and the exchanges across the system boundary are described in terms of the prices of the commodities exchanged. The justification for this lies, ultimately, in the belief in the 'perfect market' theory of price determination. The perfect market theory *assumes* that all producers and consumers have perfect knowledge of the present and future availability and production costs of all commodities. Were this the case then the conventional analysis, based on prices, would give all the information required for an investment or policy decision. However, this is not the case. Knowledge of the present is not freely available and knowledge of the future does not exist, so present prices cannot give information on the future availability or production costs of commodities.[2] Clearly any future changes in the availability or costs of production will eventually be experienced as price changes, but the point is to try to anticipate these changes. There are many factors, such as cartel formation, union action, government taxes and subsidies, which will affect future prices which cannot be anticipated by any method of analysis. However, there are also technological features of availability and production costs which are amenable to analysis.

This can be illustrated by considering an alternative way of estimating the total financial cost which involves choosing a set of basic commodities and evaluating the total quantity purchased including direct and indirect inputs.

[2] Some analysts have argued in a circle by trying to deduce information about the future from present prices. Clearly if the market were perfect, a necessary condition for prices to contain future information, then everyone would have perfect information about the future without recourse to analysis!

If the quantity of commodity i is Q_i per unit output and its price P_i then the total cost estimate, per unit output, C, is

$$C = \sum_i P_i Q_i. \tag{1}$$

This type of commodity analysis can be applied to any number of basic commodities, although care is necessary to avoid double counting.[3] Energy analysis is a type of commodity analysis since it sets out to evaluate the total quantity of one commodity, fuel, used in the production process. Energy analysis can be related to financial analysis by considering the subdivision of all the factor inputs to a production process, as shown in Fig. 1. Ultimately all the factor inputs result in a 'personal income'. However, for the purpose of this

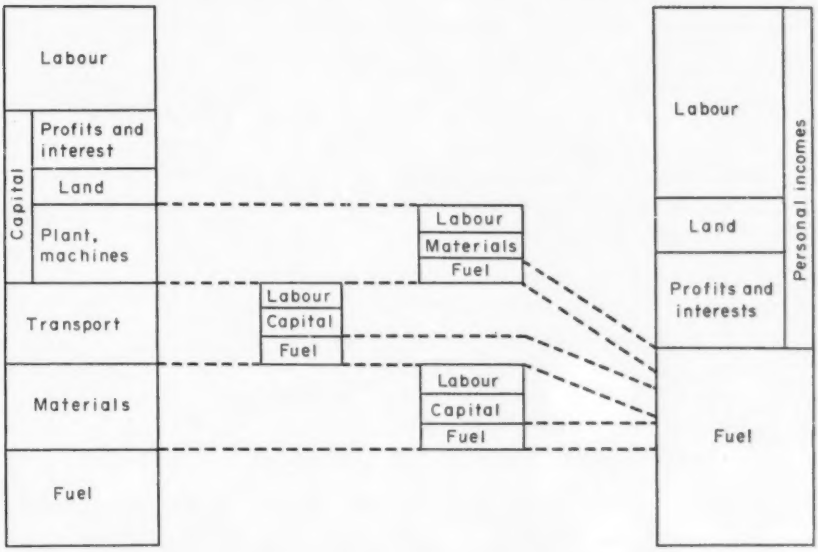


FIG. 1. *The division and subdivision of factor payments into 'payments to personal income' and 'payments for fuel'.*

analysis payments to the fuel industries are held separate and represented by a quantity of energy times a price of energy.[4] Thus, in terms of an energy analysis framework the total cost per unit of output, C, is given by

[3] Evaluation of commodity inputs can be accomplished using I/O tables, see Ref. 49. The problem is that I/O tables are usually too aggregated to be useful.

[4] The justification for not dividing fuel payments into personal incomes is that a large fraction of fuel payments is to an external cartel. Furthermore, the price of other fuels will rise towards the cost of this monopolistically controlled fuel price since there is a large degree of substitution possible between fuels.

$$C = p_e E + \sum_j P_j \qquad (2)$$

where p_e = price per unit energy
E = energy per unit output (i.e. ger of output)
P_j = personal incomes for labour, finance, land, etc.

This immediately enables the effect on costs of a change in the price of energy to be calculated as

$$\Delta C = \Delta p_e . E. \qquad (3)$$

This method of predicting the effect of fuel price rises can be extended to incorporate different rises for different fuels (e.g. Ref. 6). As explained below, energy analysis may also have a role to play in estimating the future price of fuels relative to other commodities.

4. APPLICATIONS OF ENERGY ANALYSIS

There are probably as many reasons for performing energy analyses as workers in the field. However most of the published studies fall under one of the following general headings:

1. evaluation of energy policy, energy sources, etc. [12, 15, 21, 25, 40, 42, 43];
2. effects of fuel price rises [6, 37];
3. evaluation of conservation measures [14, 22, 27, 36];
4. energy use in agriculture [29, 39, 47];
5. energy and material resources [2, 7, 13, 16, 31, 41];
6. energy use in transport [2, 26, 33, 34];
7. energy and product design, technical change etc. [1, 3, 4, 5, 23, 35, 49].

Most of these studies are concerned with the evaluation of energy requirements or energy intensities. Whilst most studies relate their findings to policy issues, only a few have any direct application. This is a reflection of the youth of the subject and the fact that a considerable data base is necessary before direct applications can arise. The most constraining factor in energy analysis at the moment is the availability of reliable data and it is in this area where most activity currently takes place. In principle, once collected this data should lead to a number of important applications. At the moment these can only be illustrated by incomplete studies, but they do show how the subject could develop into a useful policy tool.

(a) *Effect of fuel price rises*

The division of commodity inputs into payments to 'personal incomes' and to the fuel industries, as shown in Fig. 1, allows the effect of fuel price rises to be

calculated according to one of two assumptions. Either 'personal incomes' (excluding those derived from employment in the fuel industries) maintain their real purchasing power, in which case all prices rise in proportion to fuel prices, or it can be assumed that personal money incomes remain constant, in which case the price rise can be calculated using equation (3). In practice neither assumption holds, there is some loss of purchasing power, but not in direct proportion to the fuel price rise. However, for the more energy intensive industries the calculations based on equation (3) do give a good indication of which prices are likely to increase the most. Table 1 shows the range of price

TABLE 1. COMPARISON OF CALCULATED AND OBSERVED PRICE RISES FOR BUILDING MATERIALS

	Calculated rise (%)	Observed rise (Dec. 73–Dec. 74) (%)
Timber	6–7	0
Sand	7–25	20
Bricks	11–16	26
Paint	18–28	33
Mild steel	19–25	34
Cement	35–37	36
Plaster	13–15	42
P.V.C.	30–40	50

rises calculated using different estimates of energy requirements [6, 17] and the actual price rises observed from December 1973 to December 1974 [9] for a number of building materials. Except for 'plaster' the ranking of the calculated price rises agrees with the observed rises.[5] For less energy intensive products there is significantly less correlation between calculated and observed price rises.

(b) *Oil shales*

Energy analysis may have a special role to play in refining the financial analyses of future energy sources. This is perhaps best illustrated by the recent reports [44] of escalating costs of oil shale production. "In the past each time the price of oil went up, shale oil companies promptly declared that their time had finally come. . . . Now companies which once said shale oil would be profitable at $5.0, then $7.50, then $11.0 a bbl are hard pressed to name any figure at all."

According to equation (2) the costs of producing a barrel of oil from oil shales can be written as

$$C = p_e E + P \tag{4}$$

[5] Note that the lower than predicted price rise for timber could be accounted for by the very large increase in timber prices prior to December 1973 due to the increase in demand.

where P represents the sum of all personal income payments. If we accept an estimate of C = \$6/bbl when OPEC oil was \$2.5/bbl, and the present estimate that C = \$12/bbl when OPEC oil is \$10/bbl, then we can estimate E by assuming a rate of inflation for P, the personal income payments. Assuming that the price of energy (p_e) is tied to the price of OPEC oil and that in the period considered the personal income payments inflate by 20 per cent then we have

$$6 = 2{\cdot}5\,E + P$$
$$12 = 10\,E + 1{\cdot}2\,P$$

giving

$$E = 0{\cdot}69 \text{ bbls oil input/bbl oil output.}$$

Thus the price escalation could be explained by a very high energy input needed per unit energy output. The above calculation should *not* be considered as a way of calculating E since the final result is very sensitive to the cost estimates (which until recently were probably not much more than guesses) and the inflation factor assumed for personal incomes. The most reliable way of estimating E is by performing a detailed energy analysis of the process. Given an accurate estimate of E then the price of oil at which oil shales are economic is given by

$$p_e = \frac{P}{(1-E)} \text{ \$/bbl.} \qquad (5)$$

This assumes that all the energy inputs are at the prevailing oil price.

Essentially the energy analysis of energy resources points to a feedback between inputs and outputs which could be overlooked in a financial analysis. Although a financial analysis could accommodate the feedback due to the direct consumption of fuel it could not take into account all the indirect fuel consumption without resorting to an energy analysis procedure.

It should be noted that equation (5) shows that the relative prices of energy derived from different sources depends upon two factors, the payments to 'personal incomes' P, and the net energy yield, $(1-E)$. It is possible that some of the previously uneconomic energy sources now being considered as substitutes for OPEC oil could be even more expensive than anticipated due to the 'net energy' term. For example it has been estimated that uranium could be recovered from Chatanooga Shales at \$70/lb U_3O_8 [8] compared to present uranium prices of \$10/lb U_3O_8. Energy analyses of the production processes involved [12] shows the Chatanooga shale to be significantly more energy intensive than present mining with the result that the recent oil price rise only increases the price of \$10/lb uranium by 5 per cent but the \$70/lb uranium rises by 30 per cent.[6]

[6] The energy requirement for uranium from Chatanooga shales published in Ref. 12 is incorrect. Correction to the water and sulphuric acid inputs give an energy requirement of $7{\cdot}11 \times 10^6$ kWht/ton U_3O_8.

Vol. 4
1976

(c) *Long-term price trends of materials*

The theoretical minimum of energy required to mine enough ore to produce a ton of metal is inversely proportional to the ore grade. The theoretical minimum of energy is required to remove the ore from the ground and to crush the ore prior to separating the mineral particles from the dirt. Over the past 70 years the average grade of copper ores worked has declined from about 2·5% Cu to less than 1% Cu, with the U.S.A. currently exploiting ores containing less than 0·5% Cu [30, 32]. Thus, over this period the theoretical energy required to mine 1 ton of copper has increased. Over the same period there have been technical improvements in the efficiency of using fuels to perform the transport and crush ores. However, these technical improvements, particularly in the design of engines, electricity generators and crushing machines, are now more difficult to achieve since the technology is approaching a limit. These time trends, of increasing theoretical energy requirements and improvements in the efficiency of using fuels are shown schematically in Figs. 2(a) and 2(b). The actual quantity of fuel required per ton of metal at any time is the theoretical requirement divided by the efficiency of fuel utilisation. The variation deduced from Figs. 2(a) and 2(b) is shown in Fig. 2(c). This indicates that although fuel inputs to mining may have declined historically this is no guarantee that they will not rise, perhaps very rapidly, in the future. The point at which the fuel requirements start to rise very quickly is when the technological efficiency approaches its theoretical limit *and* when the average grade of ore declines at a relatively fast rate. Thus, according to this analysis the point at which material prices start to increase ought to be determinable by examining trends in technical efficiency and ore grades. Energy analysis should also be able to identify which production costs will rise most since this should depend on relative energy intensity.

There is some evidence to support this view of the long-term fuel requirements for material production. Figure 3 shows the fuel use per unit output of *all* U.S. mines [30]. It is not yet clear how much the increase in fuel requirements leads to an increase in costs, nor have there been enough energy analyses of different materials to indicate which will rise fastest or when rises are likely.

(d) *Evaluation of conservation schemes*

The main role of energy analysis in the evaluation of measures designed to conserve energy is that it considers the entire production system and incorporates both indirect and direct fuel consumption. A review of copper smelting furnaces [48] recommended an electric furnace on the grounds that it had a thermal efficiency of 61 per cent compared to oil or gas fired furnaces with efficiencies of about 27 per cent. When the efficiency of generating electricity is included in the energy analyses of the two furnaces then the electric furnace has a significantly lower performance in terms of *primary energy consumption* [10].

A more complex situation arises in considering the use of industrially gener-

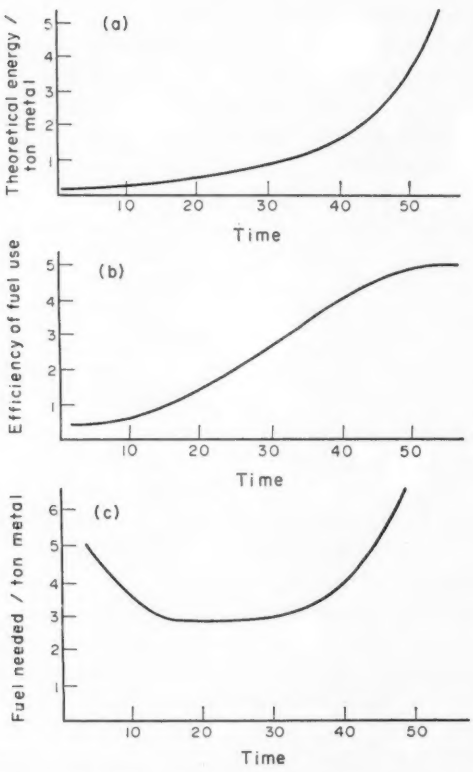


FIG. 2. (a) *The increasing theoretical energy requirement per ton of metal as the grade of ore worked decreases with time.* (b) *The change in technical efficiency of utilising fuels over the same time period.* (c) *The quantity of fuel required per ton of metal deduced from* (a) *and* (b).

ated scrap for producing new material. A steel company might conclude that the more steel it produces from scrap rather than pig iron, then the smaller will be its energy consumption per ton of steel output. This might encourage the steel company to install scrap-handling furnaces and increase the price paid for scrap. Let us assume that at the same time a car-manufacturer chooses to install a steel press which has a smaller direct fuel consumption but produces more scrap steel than an alternative press. The car manufacturer does not suffer a serious financial penalty since the price of steel scrap has been increased by the steel company. Indeed, the car-manufacturer will perceive his investment as a way of reducing costs by 'conserving energy'.

However, the net effect of these two decisions is to significantly increase the fuel consumed per motor car produced. This is obvious when the system studied includes both the steel-works and car manufacture plant, as shown in

Vol. 4
1976

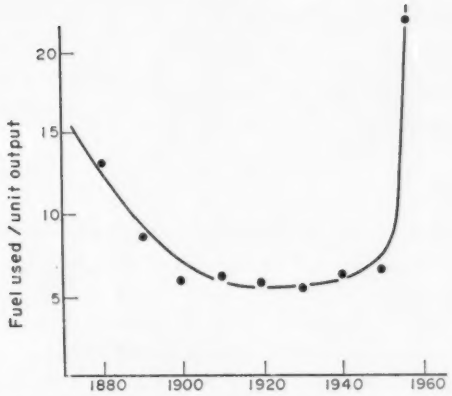


FIG. 3. *The fuel required per unit output of all U.S. mines.*

Fig. 4. If all the scrap steel used by the steel company comes from the car manufacturer then the larger the flow of scrap the larger will be the energy required per motor car. From Fig. 4 the quantity of pig iron per motor car is unaffected by the scrap cycle, but the steel furnace fuel consumption will increase as the amount of scrap increases.

This is an extreme example of how optimising part of a system for minimum direct fuel consumption could lead to an overall increase in fuel consumption. Paradoxically such situations are most likely to arise when fuel supplies are expensive and 'energy conservation' is fashionable.

At a more general level energy analysis can also identify the areas where most significant fuel savings could be made. Also, by comparing the actual energy expended in a process with the theoretical minimum energy required, energy analysis can identify areas where expenditure on research may offer significant returns in terms of energy savings. It is already apparent that many processes

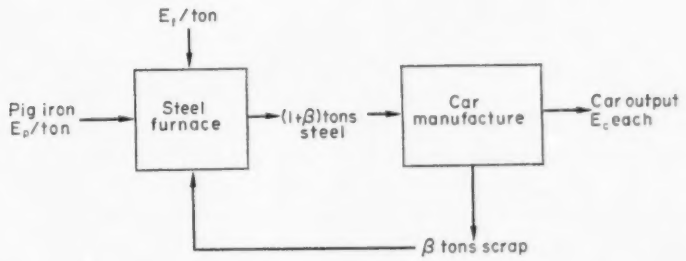


FIG. 4. *The system including both steel-works and car-production plant shows that the energy required per car is given by* $E_c = E_p + E_f (1 + \beta)$ *where* β *is the fraction of steel recycled and* E_f *the quantity of fuel per ton throughput for the furnace.*

which are dominated by a change in enthalpy (e.g. chemical processes) or by work requirements (e.g. transport) are close to their thermodynamic limit whereas processes which are dominated by a change in entropy (e.g. mineral benification, uranium enrichment) are many orders of magnitude away from their theoretical minimum requirements.

(e) *Energy policy*

Energy analysis can contribute to both the evaluation of energy resources, as explained above (b), and to the evaluation of energy demand. Until recently most estimates of future energy demand were based on the aggregate correlation between fuel use and GNP. Energy analysis relates fuel use to commodities in final demand and as such permits a disaggregated forecast which can allow for changes in the production and demand for different commodities [43]. A 3 per cent rise in GNP means that real production, and hence real purchasing power, is increased by 3 per cent. However, if the increase in production were due to an increase in demand for oil products ($\sim$2500 MJ/£) it would lead to an increase in primary fuel demand a hundred times greater than if the increase in production was in cigarette manufacture ($\sim$25 MJ/£). This emphasises the significance of patterns of production, and hence patterns of expenditure and lifestyles, on the relationship between GNP and fuel use.

This technique of estimating future fuel demand could be extended to incorporate any foreseeable changes in technological efficiency, product design or resource depletion. In addition energy analysis enables investment decisions which involve increasing fuel consumption, such as house insulation [11], to be incorporated in the projections. Although energy analysis provides a way of evaluating policy options in detail it has only been used in a few instances [18, 20].

Vol. 4
1976

5. CONCLUSIONS

Energy analysis is a new subject which may provide useful insights and data on the use of resources in an industrial system. To date most studies have been concerned with establishing basic data on the energy requirements of goods and services. This must continue to be the most important activity in the field since without reliable data there can be no application. The studies could be enlarged so as to compare processes, not just products, and evaluate both the marginal and the average energy requirements. A major flaw in most energy analyses reported to date is that they do not contain estimates of errors or reliability. This defect of presentation should be corrected at once. The reliability of the energy data and validity of analysis methods will only become clear when the same processes or technologies are evaluated by several different studies.

Any report of an energy analysis must contain a clear statement of the conventions and methods used in the study. To facilitate comparison with other workers the conventions should be based on the IFIAS guidelines [19], any departures from this or any additional assumptions should be clearly stated and related to the aims of the study. In this context it is always important to state the assumed energy content of primary inputs; the energy conversion efficiencies used and the size of the system being analysed. Without clear statements on all these points then the results of an analysis may be misinterpreted or applied in a fallacious way.

Since energy analysis provides a way of looking at the overall performance of an industrial system its results will probably be more useful for the formulation or evaluation of government policy and planning than for business planning. In this respect energy analysis provides a very versatile way of estimating energy demand and can take account of changes in technology and patterns of consumer expenditure. Energy analysis can provide good guidelines for a fuel pricing policy which will discourage the least efficient use of primary fuels.

Energy analysis can also provide information which could improve financial cost estimates for business since it gives the price elasticities and some guidance on likely price trends. These are particularly significant for the most energy intensive industries, which are those concerned with the production of fuels and materials. At the moment all these applications must await the compilation of a reliable data base. In the longer term the methods of analysis could be extended to evaluate other resource requirements of production processes, particularly 'man-hours' and 'land'. Together these resource analyses provide a framework for the evaluation of technical changes and for refining estimates of future financial costs.

ACKNOWLEDGEMENTS

The author is grateful to members of the Energy Research Group, particularly Professor M Hussey, for many useful discussions and comments on this paper.

REFERENCES

1. ATKINS PR (1973) Recycling can cut energy demand. *Engng Min. J.* **1974**(5) 69.
2. BERRY RS, FELS MF and MAKINO H (1974) A thermodynamic valuation of resource use: making automobiles and other processes. In *Energy: Demand Conservation and Institutional Problems.* (Ed. MACRAKIS MS) p. 499. MIT Press, Cambridge, Mass.
3. BERRY RS and FELS MF (1973) The energy cost of automobiles. *Bull. At. Sci.* (*Sci. Pub. Aff.*), **10,** 11.
4. BERRY RS and MAKINO H (1974) Energy thrift in packaging and marketing. *Technology Rev.* **76,** 33.
5. BERRY RS and LONG TV (1975) An international comparison of polymers and their alternatives. *Energy Policy* **3,** (in press).

6. BRAMSON MJ and VINES DA (1975) Price propagation in an I/O model: determining the implication of higher energy costs for industrial processes. *Proc. Conf. Understanding Energy Systems (session C), London, April* 1975 (also to be NEDO publication).
7. BRAVARD JC, FLORA HB and PORTAL (1972) *Energy expenditures associated with the production and recycle of metals.* Rpt. No. ORNL-NSF-EP-24 (Oak Ridge Nat. Lab., Tenn.).
8. BIENIESKI CL, PERSSE FH and BRAUCH EF (1971) *Availability of Uranium at Various Prices from Resources in the U.S.* U.S. Bureau of Mines. Circ. 8501. U.S. Govt. Printing Office.
9. *Building* editorial staff (1975) Monthly cost information file. *Building* **3**
10. CHAPMAN PF (1974) Energy costs—a review of methods. *Energy Policy* **2**(2), 91.
11. CHAPMAN PF (1975) Conventions, methods and implications of energy analysis. In *Proc. Conf. Understanding Energy Systems, London, April* 1975.
12. (a) CHAPMAN PF (1974) The ins and outs of nuclear power. *N. Sci.* **64**, 866.
    (b) CHAPMAN PF and MORTIMER ND (1974) *Energy inputs and outputs of nuclear power stations.* Open University Research Rpt. ERG 005.
13. CHAPMAN PF (1974) Energy costs of producing copper and aluminium from primary sources. *Met. Mater.* **8**(2), 107.
14. CHAPMAN PF (1974) Energy conservation and recycling of copper and aluminium. *Metals Mater.* **8**(6), 311.
15. CHAPMAN PF, LEACH G and SLESSER M. (1974) The energy cost of fuels. *Energy Policy* **2**(3), 231.
16. CHAPMAN PF (1975) The energy cost of materials. *Energy Policy* **3**(1), 47.
17. CHAPMAN PF and MORTIMER ND (1975) *Energy analysis of the Census of Production* 1968. Open University Research Rpt. ERG 006
18. CHAPMAN PF (1975) *A Fuel's Paradise–Energy Options for Britain.* Penguin Books, London.
19. *Energy Analysis* (1974) Workshop Rpt. No. 6. IFIAS (Nobel House, Sturegatan 14) Stockholm.
20. *Exploring Energy Choices* (1974) Publ. Energy Policy Project (Ford Foundation).
21. FOLK H and HANNON B (1974) An energy pollution and employment policy model. In *Energy.* (Ed. MACRAKIS MS), p. 159. MIT Press, Cambridge, Mass.
22. HANNON B (1974) Options for energy conservation. *Technology Rev.* **76**, 25.
23. HANNON B (1972) Bottles cans and energy. *Environment* **14**(2), 11.
24. HERENDEEN RA (1974) Use of input–output analysis to determine the energy cost of goods and services. In *Energy.* (Ed. MACRAKIS MS), p. 141. MIT Press, Cambridge, Mass.
25. HILL KM and WALFORD FJ (1975) Nuclear aspects of energy accounting. In *Proc. Conf. Understanding Energy Systems, London, April,* 1975.
26. HIRST E (1972) *Energy consumption for transportation in the U.S.* ORNL-NSF-EP-15 (Oak Ridge Nat. Lab. Rpt.).
27. HIRST E and MOYERS JC (1973) Efficiency of energy use in the U.S. *Science, N.Y.* **179** (4080), 1229.
28. KENWARD M (1975) How to cook the energy accounts. *N. Sci.* **66**(946), 205.
29. LEACH G (1973) The energy costs of food production. In *The Man–Food Equation.* (Ed. BOURNE A). Academic Press, London.
30. LOVERING TS (1969) Mineral resources from the land. In *Resources and Man.* Freeman, London.
31. MAKHIJANI AB and LICHTENBERG AJ (1972) Energy and well-being. *Environment* **14**(5) 10.
32. *Mineral Facts and Problems* (1970) U.S. Bur. Mines Circ. 650, U.S. Govt. Printing Office.
33. MORTIMER ND (1974) *Gross energy requirements of marine transport.* Open University Research Rpt. ERG 007.
34. MORTIMER ND (1975) *Gross energy requirements of air transport.* Open University Research Rpt. ERG 008.
35. MOYERS J (1972) *Room air conditioners as energy consumers.* Rpt. ORNL-NSF-EP-59 (Oak Ridge Nat. Lab. Tenn.).
36. N.E.D.O. (1974) *Energy Conservation in the U.K.* HMSO, London.
37. N.E.D.O. (1974) *The Increased Cost of Energy—Implications for U.K. Industry.* HMSO, London.

38. NILSSON S (1974) Energy analysis—a more sensitive instrument for determining costs of goods and services. *Ambio* **3**(6), 222.
39. PIMENTEL D *et al* (1973) Food production and the energy crisis. *Science, N.Y.* **182,** 443 (see also *Science, N.Y.* **184,** 135, 307).
40. PRICE J (1974) *Dynamic Energy Analysis of Nuclear Power.* Earth Resources Ltd. London.
41. ROBERTS F (1974) Energy consumption in the production of materials. *Metals Mater.* **8**(4), 205.
42. ROBERTS P (1973) Models of the future. *Omega* **5**(5), 592.
43. ROBERTS P and OUTRAM VE (1974) *A Method of Projecting Energy Demand in the U.K.* Dept. Environment, London.
44. (1975) Shale oils high risk future. *Bus. Wk,* April 28, 87.
45. SMITH H (1969) Cumulative energy requirements of some products of the chemical industry. *Transactions* (Wld. Pwr. Conf.) **18.**
46. SLESSER M (1975) Accounting for energy. *Nature, Lond.* **254,** 170.
47. SLESSER M (1973) Energy subsidy as a criterion in food policy planning. *J. Sci. Fd Agric.* **24,** 1193.
48. TRIELHARD DG (1973) Copper state of the art. *Chem. Engng* (special review April issue).
49. WRIGHT DJ (1975) Primary resource requirements of commodities *Appl. Econ.* **7**(1), 31.
50. WRIGHT DJ (1974) Calculating energy requirements of commodities from the I/O table. *Energy Policy* **2**(4), 307.

ADDRESS FOR CORRESPONDENCE: *Dr.P Chapman, Director, Energy Research, The Open University, Walton Hall, Milton Keynes, MK7 6AA*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

A Framework for Optimizing Managerial Decision

SB LITTAUER [1]
TM YEGULALP [2]
GK ZAHARIEV [3]

(*Received March 1975; in revised form July 1975*)

There have been numerous attempts to analyze and formulate management problems within the general framework of Management Science and Operations Research, so as to help managers in their decision making problems. This paper focuses on the fundamental aspects of management decision making and introduces a conceptual framework in which problems of management and its different levels of functioning are systematically identified. Management problems are considered within two major categories: Enterprise and Operations problems. In spite of the difficulties in defining a clear-cut boundary between Enterprise and Operations problems, such a broad classification leads to three basic levels of management functioning, namely: Policy making, coordinating and implementation. The universal relationship among these three basic levels of functioning provides a basic functional element called the Management Triad. In contrast to classical higherarchial structure, triad structure provides a methodology to analyze relationships among different levels of management functions in terms of a given decision problem. It is demonstrated that for different decision making problems individuals can shift from one end of the triad to another and perform a different management function.

A quantitative approach has been introduced for operations management decision making by adapting Shewhart's control chart philosophy which in turn provides for quasi-automated decisions for corrective action and also for filtering significant signals of change to appropriate management functioning levels.

INTRODUCTORY REMARKS

THERE are many perspectives from which management has been viewed. With the development of quantitative techniques in decision making, in particular under the influence of Operations Research and Management Science,

[1] Professor Emeritus Operations Research, School of Engineering and Applied Science, Columbia University, New York.

[2] Associate Professor of Mining, School of Engineering and Applied Science, Columbia University.

[3] Now Associated with: Mining Department, Higher Institute of Mining and Geology, Sofia, Bulgaria. Formerly associated with Krumb School of Mines, School of Engineering and Applied Science, Columbia University.

quantitative management has been the aim of investigators in the field. It is in keeping with the spirit of modern management concepts that this paper is presented.

We seek to establish within management a methodology from the perspective of the following aims:

1. Establishment of a general conception and characterization of the functioning of management.
2. Providing quantitative criteria for operations management decision making.
3. Providing for a quasi-automated management decision process.
4. Developing an action-oriented management information system structure.
5. A simplification of the conceptualization of management.

Some years ago Merrill M. Flood [1] suggested that *management*, or better, the *management function*, is *conscious goal-directed decision making*. While this view may not be striking, it does remove much of the mystique which seems to surround the manager who is charged with making decisions under circumstances which are perplexing. Yet the simple concept of management offered by Flood, which has general and broad applicability, can provide the basis for specifying operationally the means for achieving and evaluating the effectiveness of management performance. From this perspective we develop principles for designing adaptive management information systems, and adaptive managerial decision procedures.

CHARACTERIZATION OF MANAGEMENT IN TERMS OF MANAGEMENT PROBLEMS AND FUNCTIONS

It has been found helpful in devising an approach for analysis and solution of Management problems to consider two broad classes of problems which differ markedly. Furthermore, it is a fact that most of the actual problem solving, as well as most of the literature of Operations Research and Management Science, has been devoted to one class of these problems. We explore this classification together with the task of coordinating the two principal categories. This coordinative function is examined as a management category developed below.

A. Operations problems

Much managerial effort is exerted in keeping various operations of an industrial business enterprise at a certain desired level of functioning. Most techniques of Operations Research are effective on operations problems.

Characterizing operations and operations problems is a prerequisite for effective managerial decision in this category. We shall characterize operations problems and enterprise problems by listing respectively the properties which identify them. Later we will discuss the levels of management functioning which we feel provide for communication at various levels of decision that can be treated at these levels in terms of those of their characteristics which determine techniques of decision procedures essential to effective decision.

Operations and Operations Problems:

a. Are ongoing repetitive activities.

b. Are subject to the influence of many internal causes in contrast to environmental causes; internal causes can be manipulated by the persons responsible for the operations.

c. Have outcomes whose variation is measureable; associated probabilities can be estimated, variations can be made statistically stable, and thus future outcomes can be probabilistically forecast with estimable measures of error.

d. Incur penalties as a result of variations beyond control limits but the penalties therefore are relatively not high; these penalties are definitely not catastrophic.

e. Are reversible in the sense that:
   i. the presence of assignable causes can be detected;
   ii. these causes can be eliminated, so that the operation can be brought back to its intended state.

f. Are susceptible to interaction with operators, and observers taking measurements on the operation. For example, worker performance may be biased by worker reaction to the observational process; similarly, corrective action may introduce deviations from specified parameters.

Examples of operations are plentiful. Every production process is an operation. The economic control of these processes is most effectively achieved when their characterization is appreciated. Then the Operations Research–Management Science approach and the specific techniques of Statistical Quality Control can lead systematically to greater attainment of management objectives with respect to operations.

Among the variety of types of operations problems are: determining the mixture of items to be manufactured by way of allocation of scarce resources (production facilities) so as to maximize the profit therefrom; assigning the appropriate number of repairmen so as to minimize loss from down time; a highway toll collection system provides a number of operations problems; maintaining an airline reservation system in effective functioning; maintaining

an inventory policy system; scheduling job-shop operations; managing a typing pool. The variety of operations is great, and indeed, one very important level of management is concerned with the control of operations. We repeat again that with an appropriate approach, the techniques of Operations Research and the Management Sciences are genuinely effective in the resolution of *operations problems.*

B. Enterprise problems

In contrast to operations, there is the *enterprise* itself. An enterprise, of course, has many aspects other than those of operations, and there are many operations within the enterprise. The interrelatedness of certain groups of these operations creates many problems apart from those specific to the individual operations considered separately. Some of these interrelations among operations may not create serious issues affecting the wellbeing of the whole enterprise; a problem arising in connection with a particular operation can affect other operations, but rarely does it affect the whole corporate entity. For example, a poorly conceived inventory policy for the manufacture of a particular product may lead to losses owing to slowing of production of that item, and thus to some reduction of profit of a subdivision of the corporation. It is not likely, however, to cause serious weakening of the corporate position. Again, poor quality control on some product may both raise production costs and reduce sales without seriously damaging the division's or the corporation's future. This is not to say that ineffective management of operations does not entail serious consequences, nor that such ineffectiveness can be permitted to occur without corrective action. But as has been stated above, the techniques for operations management decision that have been developed provide for corrective action which restores an operation to healthy functioning before losses go beyond safe economic bounds. Furthermore, the approach to be developed here provides decision procedures on the control of operations so as to prevent the occurrence of inventory losses, quality losses and the like. We already have within the body of Management Science and Operations Research a range of techniques for operations management, which enable one to design inventory policies, production planning and scheduling procedures, resource allocation procedures, quality control policies, and the like, so as to attain the greatest output for a given input. And in the nature of the methodology and techniques to be developed here, decision on the control of operations are subject to continuing corrective adjustment. This aspect of what we call Operations Management will be discussed below.

In contrast to this there are decisions confronting management which can affect the whole enterprise seriously. Wrong decisions on plant location, new product lines and even labor policy can set back the growth and profit position of a company for many years and possibly lead to bankruptcy. We call this

Vol. 4
1976

type of problem an *enterprise problem*. We characterize enterprise problems as follows:

Enterprise problems:

a. Are non-repetitive; entail essentially 'one-shot' major decisions.

b. Entail non-reversible decisions.

c. Are subject to a great many variables which are largely environmental (external) and highly variable. Furthermore external variables cannot be manipulated.

d. Entail difficulties in the measurement of the effects of these external variables.

e. Leave the question of statistical stability in doubt; data rarely comes from repetitive sources; probabilities cannot readily be estimated.

f. Entail very high, sometimes catastrophic penalties for wrong decisions.

The differences between enterprise and operations problems derive from the approach to enterprise problems, which is rather different from that approach which is effective on operations problems. Although there are many treatises on market analysis, forecasting, plant location, corporate merging, diversification, capital investments, long range planning, strategic planning and other enterprise problems, there are no established approaches nor systems of working techniques that can be directly applied as in the case of many operations problems. This does not mean that we cannot undertake systematic investigations of enterprise problems using the whole gamut of Operations Research and Management Science techniques coupled with managerial ingenuity. It does mean, however, that we must exercise caution in planning such investigations, being alert to the pitfalls inherent in the large-scale enterprise problems. Enterprise problems may consist at times of many operations problems whose total effect differs from the 'sum of the individual effects'. The relationships between the variables of operations problems and the enterprise problem must be very well understood; validity of certain assumptions essential for the application of some techniques of Operations Research and Management Science must be seriously questioned when considering enterprise problems. Furthermore, if all subsystems of a large system function at their optimal levels, it does not necessarily follow that the system itself is functioning at its optimum. These are the major areas where the pitfalls of a large-scale enterprise problem are usually encountered.

THE HUMAN LINKED SERVO-CONTROL

Effective decision making requires the interrelating of Action-Outcome and Inference in such a way that, given a situation in which a question arises for

management decision, one can infer from the observation of an outcome that has occurred, what action or set of actions caused that outcome. We are using the word 'outcome' above as the word 'effect' is used in the relation cause and effect. In this sense our use of the word 'action' is that of 'cause'. We prefer action-outcome as a more primitive expression than cause and effect. We use the word 'inference' in the sense that, having observed an outcome we go through a mental process which suggests an action (or set of actions) from which the observed outcome can be expected to follow. We believe that if we repeat this action, the same outcome will occur. We are not concerned with the fact that having observed the outcome of a given action an inference can be made by almost anyone. Such an arbitrary process may not relate action and outcome verifiably. What is of concern is the possible existence of some methodology comprised of techniques which can be used by management to increase the degree of effectiveness of their decision making, of their making inferences relating action-outcome. Towards this goal we propose to introduce a feedback procedure which can provide for continuing correction of a process under consideration. In effect we seek an automated servo-control of operations, in essence a mode of automated managerial decision. What actually can be achieved is more properly a form of quasi-automated managerial control of operations. In the light of our classification of the management function this objective is quasi-automated operations management which by its nature lends itself to computerization. For completeness we give some elements of the procedure which we call 'Human-Linked Servo-Control'.

Consider Fig. 1, where time order is represented on the horizontal axis and some property of a given process is represented on the vertical axis. In essence what is shown in this figure is an $\bar{X}$ control chart with 25 averages taken at some regular time intervals and plotted in time order. We aim to attain the most economic production of an item, which requires a balance between total output and satisfactory output. This can be obtained by monitoring the process so as to take corrective action whenever necessary. And of course, it is necessary to do the monitoring economically [3]. The Shewhart statistical control chart has been found to be one of the most effective tools for attaining *economic control of a manufactured product*. We view this function as a principal industrial responsibility which exemplifies possibly 'automated' or 'quasi-automated' managerial decision triggering.

The chart is used in the following way: the evidence of, say, the 20 preceding sample averages provides $\bar{\bar{X}}$, the grand mean, an estimate of the 'long-run' average value that the process will yield. This value $\bar{\bar{X}}$ will agree, hopefully, with the nominal value that has been specified. There will, however, be variations as is evident on the control chart. A sample average such as $\bar{X}_{21}$ (the 21st sample average) which lies beyond one of the bounds is a signal that there may have been a change in the causal system governing the process. Here we recognize that variation outside the calculated bounds (UCL and LCL) signals the

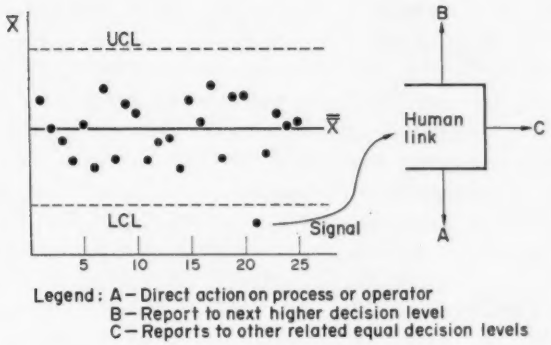


FIG. 1. *Human-linked servo control model.*

'probable' presence of a cause that is disturbing the process. For centuries we depended largely on our intuition in the control of production processes. Today we can utilize the control chart to guide us in decision making. One must bear in mind that more than one variable may be affected by the same cause. A decision process may require the use of several control charts on different characteristics related to the same process.

In our simple example an outage calls for the operator to take action: he himself may investigate the situation or he may call the supervisor to seek his intervention.

Now for the managerial aspect: actually decision on a process of any kind is in essence a managerial decision. In its broadest aspects, however, the basic decision requirements of operators and supervisors feed into decisions of broader scope concerning personnel performance, machine performance, section performance, office staff performance, salesman performance, foreman performance, and so on up the hierarchy of decision makers. As well one can treat cash flow, profit performance, advertising effectiveness, inventory variation and other factors contributing to section, plant, division and corporate performance in achieving corporate goals. The supervisor is commonly referred to as 'first-line management'. He is concerned primarily with operations decisions; the action alternatives include instructing the operator, investigating causes or requesting from one's superior analysis of causes for the 'outage', adjusting the process and performing other incidental functions essential to maintaining the process in an economic state of operation. By economic state we mean that operating condition which yields the greatest net return from the process: sales value of output less all related costs. Without going into details as to which costs are chargeable to the process we can agree that economic manufacture must strike a balance (among other things) between cost of increasing product value, by increasing the 'quality' of the product, and costs of quality control [3]. Hence the decision making function of even first-line management is important

and it is even more important that upper levels of management be fully aware of and in control of the degree of effectiveness of first-line management.

Now you may ask how the foregoing discussion bears on human-linked servo-control of processes. Bearing in mind that the operator or supervisor (or both) take direct action to correct the process upon getting the signal of an 'outage', a sample average outside UCL or LCL, we have the equivalent of feedback in an automatic control mechanism. Common examples are found in steering mechanisms used on ships and aircraft. In the case of a production process, the feedback comes from the action of the operator/supervisor in taking corrective action on the process; or, perhaps the supervisor taking action on the operator. The decision maker using control chart signals acts as a 'controller'. The control limits act like the potential levels in the electrical servo-control system, such that when that 'barrier' is exceeded a counterforce is introduced to return the system to the desired level of functioning. The control chart-decision maker combination forms a link which initiates inverse feedback with corrective action. Hence the name: Human-Linked Servo-Control for this concept of management decision making.

There are many aspects of industry operation to which the H-LS-C (Human-Linked Servo-Control) aspect of managerial decision making is especially well adapted. Among these are: labor productivity, salesman productivity, incidence of accident injury phenomena, absenteeism, analysis of customer complaints, account auditing, production costs, transportation efficiency (fuel consumption, tire replacement, repair costs), machine breakdown, worker time allowances. Throughout our discussion of the H-LS-C aspect of management we have restricted ourselves mainly to operations and related managerial decision making. The role of H-LS-C in enterprise and related managerial decision making will be more clearly understood when it is considered within the following proposed framework:

A conceptual framework for diagnosis of causes and corrective action:

1. Continuing, systematic observation of the system.
2. Recognizing or calculating or inferring the regular or irregular character or behavior of the system.
3. Establishing criteria for action when the predetermined or inferred regular behavior is changed.
4. Devising systematic procedures for identifying the 'cause' of the observed change.
5. Devising a continuing self-corrective mechanism for restoring the desired mode of behavior.

LEVELS OF MANAGEMENT FUNCTIONING

The two categories of problems with which management deals give rise to two levels of management functioning, namely: *Implementation Level Manage-*

ment and *Policy Level Management*. At both these levels of functioning, managers individually need the following five qualities: intuition—a genuine flair for policy decision making; imagination—capacity for innovation; knowledge of the business—including techniques of managerial control; managerial drive—capacity for communication, cooperation and command; courage and a sense of responsibility. At the two levels different proportions of these qualities are called into play, with all of them required for the most responsible tasks at the policy level. Yet it must be borne in mind that top managerial talent often develops through the ranks, so that even at lower implementation levels these talents must be displayed.

In order to maintain a continuity of thought, as we develop the concept of levels of management functioning, we wish to introduce the concept of 'organizedness'. There is much written about organisation without explicitly defining what is meant by the word. We have been puzzled by what we have read and have found very little consensus among different writers views. There seems to be too little in the way of general theory or in expression of concepts operationally defined. This is to be expected of a concept like that of organization which is broad and therefore elusive. There are some general notions of organization, such as line and staff, centralized and decentralized, product and functional, as well as specific notions such as 'span of control' and the assignment of certain control responsibilities to line and staff. While each of these notions and their alternatives are well-supported by relatively sound arguments, there does not seem to be any unified conception of organization nor a view of organization theory which leads to a state of *organizedness*; that is to say, *unified functioning of the total system*. From observation of various corporations it seems that building an organization structure around key individuals is a frequent practice, which practice does not necessarily lead to a state of organizedness. We hope to link the concepts of Implementation Level Management and Policy Level Management with that of the concept of *Coordinative Level Management* and so to unify the concept of effective management functioning with that concept of organization built upon a *Three Level Management Perspective*, adapting to its purpose the *Human Linked Servo-Control concept*.

We have remarked that organization structure was not a guarantee of organized functioning. Nevertheless, organization structure can contribute to or it can interfere with management functioning. We build our concept of structure on the base of three level management functioning which can in turn be geared to unified corporate functioning.

For the moment, we set down the three levels by name:

1. Policy level management
2. Coordinative level management
3. Implementation level management

and consider as an example a simple organization. It makes two products, which it markets directly to dealers. The bread and butter comes from selling these two products; the appeal of these products is their 'quality' and their selling price. Hence the manufacturing operations are basic to the survival of the company. For operations there are required a number of machine operators, two production supervisors, and one foreman who functions as a production manager. Also needed are a design engineer, who is also responsible for other engineering functions than design (including quality control) and who reports to the foreman; a salesman who is also the sales manager, and who assists the top executive officer in promotion and public relations. The top executive officer handles the finance, public relations, promotion, personnel, purchasing and all corporate policy. He has one staff assistant. Legal matters are handled by an outside attorney on a retainer basis. It is clear that most of the tasks performed by the chief executive officer indicate that he deals mainly with enterprise problems and we can identify his functioning as that of policy level management. It is not very clear, however, if one looks strictly from the hierarchical point of view, at which levels of management the others function. A careful look at the functions and possible decision making problems indicates that this three level structure can represent actual management decision practice within this organization. The three level relation is not human-assigned, as is the hierarchical structure. For example, in launching a new product (Figs. 2 and 3) final decisions are made by the top executive officer who certainly, in this case, functions at the policy making level; and the staff assistant, the design engineer and the production manager function in a coordinative capacity. The salesman operates at the implementation level in marketing and advertising decision

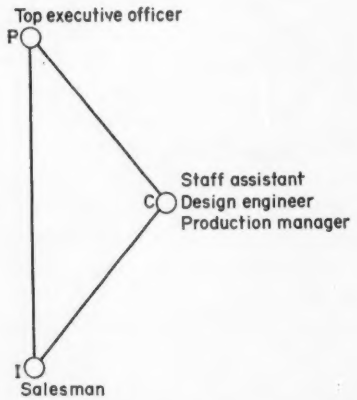


FIG. 2. *The three levels of management functioning in advertising and marketing decision making.* P *indicates the policy level,* C *indicates the coordinating level, and* I *indicates the implementation level of management functioning.*

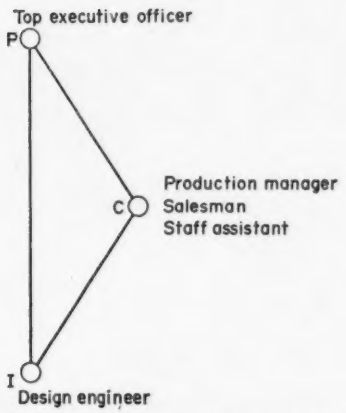


FIG. 3. *The three levels of management functioning in product design and quality control decision making.* P *indicates the policy level,* C *indicates the coordinating level and* I *indicates the implementation level of management functioning.*

making (Fig. 2), as does the production manager in manufacturing, and the design engineer in product design and quality control operations related to this new product (Fig. 3).

We see from this simple example that the three level structure is identifiable with respect to the decision problem in question. We can say further that even in this simple case one can build many interrelated three level structures (triads) for different management decision problems. The connecting line defines the information flow. So far we have not dealt with the direction and intensity of this information flow, which may well be the most important indicator of the degree of *unified functioning*. It is very revealing to examine the structure of management triads of an organization (even this simple one) for its specific problems, and to identify on this structure the directions and the intensity of information flow.

We have mentioned earlier that this company produces and markets two products. One of its most important management problems is to maintain a 'certain' quality level of its product output. This involves a decision making mechanism which we shall try to identify with management triads. In Fig. 4 we represent the triad structure related to quality control. Machine operators under production supervisors 1 and 2 perform routine observations of 'quality', functioning at the implementation level. They may be allowed to take corrective action required for maintaining the specified level; such action may entail, for example, machine adjustments, modifying worker performance and the like. One machine operator in each group (1 and 2) is charged with assisting his supervisor in addition to his normal operating tasks. In this way he acts as a coordinator of activities between the supervisor and the other operators in his group. For this function he is, of course, identified at the coordinating level.

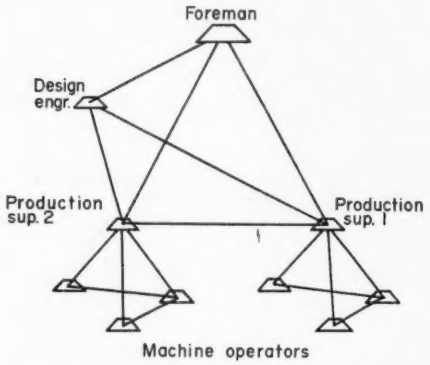


FIG. 4. *A simplified triad structure for quality control decision making.*

The production supervisor is responsible, among other things, for the control of 'quality' of the product that his group manufactures. Thus he is in the position, from the observations and feedback he gets from his operators, to make decisions on adjusting and/or changing certain inputs to maintain control of product quality. Note that his decisions are also limited. He may not be authorized to replace a particular production machine, but he may be allowed to make decisions on refitting or overhauling a machine from which the output fails to meet production quality standards. In this decision task, he is operating at a policy making level with respect to the operators.

We have mentioned earlier that production supervisors report to the foreman. In case of a major decision on changing a control procedure, or introducing new measures of control, the decision is to be made by the foreman; in this situation the design engineer is functioning at the coordinative level, while the production supervisor acts at the implementation level. The design engineer's coordinative function is important when we consider him as the one who evaluates the quality control reports of supervisors and 'translates' them into actionable signals for the foreman who initiates the decisions for corrective action.

This kind of analysis of each function leads to the construction of several triad structures, through which one can follow the information flow in terms of its directions and densities on each connecting link. We consider this as a systematic way of analyzing an existing organization or evaluating alternative ways of organizing a system for a given purpose.

One can now consider the quantitative aspects of this method of analysis. We see one major quantifiable aspect, namely that of information flow density. This reflects in a way how heavily the two nodes of a triad are dependent on each other. The density can be expressed as number of daily oral contacts, number of periodic reports in a given time interval, number of action orders per given period of activity, etc. We may also see that an incoming information

Vol. 4
1976

flow density at a node is much higher than the outgoing information flow density. This property can be translated into a 'filtering' function of that level of management (Fig. 5). To complete the quantification of this type of analysis we may identify these filters with their per cent reduction capability of the information density. One must, however, not ignore the qualitative aspects of this filtering function. Reduction rate in the outgoing information must not contribute to an information loss, but it should allow only relevant and actionable signals to penetrate higher levels of the management triad structure.

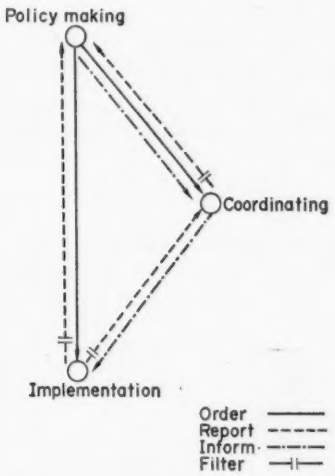


FIG. 5. *The three levels of management functioning with symbolic representation of information flow for management decision making.*

Now we can relate these ideas to the concept of H-LS-C. What H-LS-C does in a management decision system is in fact similar to a servo-control mechanism, but here action signals are generated and initiated by a human link. The incoming information is essential for the construction of a 'control-chart' or control-chart-like evaluation mechanism, where pertinent variables are observed in a continuing fashion. Outages or deviations which call for corrective action are detected and a decision for corrective action is made. Thus, in order to make such a system function properly the characteristics of filtering of such input must be selected so as to provide relevant information for H-LS-C mechanisms and also prevent dilution by irrelevant information, which we may call 'noise'.

We now have the seeds of a meaningful management information system. The task of putting these elements into their places for producing a unified functioning within the framework proposed here can be accomplished systematically. However, the resulting system designs must then be amenable to quantitative comparison for selection of the 'best' alternative.

We have tried to evolve a form of organization structure built on the framework of the three level concept of management functioning. We proposed further to adapt the H-LS-C concept to our management and organization structure concepts so as to create within a corporation a quasi-automated information system. The full name of such a system might well be: *A Quasi-automated Action-Oriented Management Information System.*

REFERENCES

1. FLOOD M (1958) private communication to SB Littauer.
2. GIBRA IN (1971) Economically optimal determination of the parameters of $\bar{X}$-control chart, *Mgmt Sci.* **17**(9).
3. SHEWHART WA (1939) *Statistical Method from the Viewpoint of Quality Control.* (Ed. EDWARDS DEMING W). Graduate School U.S. Department of Agriculture, Washington, D.C.

ADDRESS FOR CORRESPONDENCE: *Professor SB Littauer: Henry Krumb School of Mines, Seeley W. Mudd Building, Columbia University in the City of New York, NY 10027, USA.*

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

A Guided Tour of Recent Practical Advances in Integer Linear Programming

AM GEOFFRION

(*Received September 1974; in revised form May 1975*)

The field of integer programming has been an extremely active one in recent years. Only a small fraction of the contributions, however, are of demonstrable *practical* value to practitioners or software designers within the near future. These are the contributions which this paper will attempt to survey. The selections are made on a subjective basis from the published and unpublished literature appearing over the last few years.[1]

1. LINEAR PROGRAMMING ADVANCES USEFUL FOR INTEGER PROGRAMMING

SINCE linear programming (LP) provides the main computational horsepower for the leading approaches to integer linear programming (IP) (especially branch-and-bound), almost any computational advance in LP is automatically beneficial to IP. The well-known technique for generalized upper bounding (GUB), for instance, has proven valuable in IP because so many IP models contain constraints of this form. The same will probably be true of the new generation of in-core LP optimizers.

This point applies particularly to the field of large-scale linear programming for the obvious reasons and also, as indicated in more detail below, because integer programs of moderate size can enlarge impressively when seemingly redundant constraints are added for the purpose of obtaining stronger bounds. Tomlin [39] has presented a brief but incisive review of significant recent developments in large-scale LP. Another excellent review, more general in nature and with more emphasis on practical software considerations, is provided by White [40]. More recently, improved representations of the inverse have been found that permit substantially reduced pivot times and storage requirements for various specially structured large-scale problems [21, 26, 43].

Computational techniques for solving transportation and network problems

[1] Earlier algorithmic progress was surveyed in Ref. [15].

have advanced markedly. The use of better information structures has led to greatly improved specialized implementations of the primal simplex method for such problems. Several new implementations have proven clearly superior to the venerable out-of-kilter method [5, 18, 20, 23, 29]. It appears likely at this time that some of the updating techniques used in these incredibly efficient implementations admit algebraic generalizations to quite general non-network LP structures. This could have a major influence on the future of large-scale linear programming technology. In the meantime, and in any event, these network implementations can be used as the heart of branch-and-bound algorithms for a wide variety of integer programming problems that can be viewed in terms of networks.

The recent development of interactive implementations of full scale mathematical programming systems is of potentially great significance [7, 19, 25]. It should be obvious that the conveniences possible with an interactive implementation can enhance a user's productivity and overall effectiveness in almost every phase of his activities—data management, modeling, problem modification, run control, solution access and analysis, and algorithmic experimentation. The obstacles to truly interactive problem solving can be substantially diminished, a particularly important advantage in view of the fact that completely automated solution procedures for most complex problems do not exist now and never will. These advantages are especially valuable in the context of integer programming, where the principles of effective modeling and algorithmic tactics are much less routine and more in need of experimentation and closely supervised control than for linear programming. A successful practical application is described by Thomas *et al.* [38].

2. MODELING PRINCIPLES

The computational tractability of any given IP application is strongly dependent on both the *content* (assumptions) of the model and the way in which the model is *represented* mathematically (the distinction here is important). It is essential to recognize that some of the guiding principles from linear programming can be downright dangerous if applied absent-mindedly to integer programs.

Modelers should be aware that integer programming is susceptible to a pathology that rarely arises in linear programming: possible discontinuity of the optimal value as a function of constraint coefficients. If a small change in some 'soft' coefficient leads to a sudden incommensurately large decrease in the optimal cost, the analyst will very likely want to revise this coefficient or perhaps even revise the model so as to treat this coefficient as an explicit decision variable. This pathology is probably much more frequent than is commonly realized. Surprisingly, very few papers have ever addressed this topic seriously. Williams' [41] paper was among the first. Some theoretical results regarding continuity

in the IP context are established and penalty functions are recommended as a remedy for constraints whose data are likely to allow troublesome discontinuities. More comprehensive results are given by Radke [33]. Armstrong and Sinha [1] have presented an impressive modeling lesson very closely related to the discontinuity pathology. Also, Geoffrion and Graves [14] have found that minor adjustment to the problem statement can greatly mitigate the bogey of discontinuity.

Let us turn now to the topic of model representation for computational effectiveness. Discussion will largely be limited to the context of LP-based branch-and-bound, as virtually all commercially available IP software is of this type. The main lesson of Williams [42] is an important one: one should examine the various possible mathematical representations of a model which are equivalent in a logical sense and select the one which seems likely to give the tightest bound when relaxed in the usual way to an ordinary LP. The reason is that better bounds imply less need for branching, thereby shifting the balance of work to the relatively more efficient machinery of linear programming (as opposed to enumeration). Williams [42] gives five specific examples to illustrate various ways of achieving 'equivalent' formulations yielding better LP bounds. His first example shows that constraints can sometimes be made tighter without encroaching on the region of integer feasibility by adjusting coefficients systematically to exploit the integral nature of some of the variables. This point is also made forcefully by Sommer [37]. I have observed the value of this simple idea on several occasions; e.g. in the context of avoiding unrealistically large capacity limits in facility location problems.

Williams' other examples illustrate instances in which new constraints can be added that are redundant in an IP sense but are not so for the associated LP relaxation. In the second example these constraints can be discovered by graphically examining two- or three-dimensional components of the problem. For the remaining examples they can be discovered by 'disaggregating' existing constraints, as by writing $x_1 \leqslant x_4$, $x_2 \leqslant x_4$, and $x_3 \leqslant x_4$ instead of $x_1 + x_2 + x_3 \leqslant 3x_4$ when the variables are 0–1. A technique very close in spirit to constraint disaggregation, especially for problems with a preponderance of 0–1 variables, is the determination of simple logical relationships between 0–1 variables which are not already expressed by correspondingly simple constraints; e.g. if $x_1 = 1$ implies $x_2 = 1$, then one may introduce $x_2 \geqslant x_1$. Hammer and Nguyen [22] have described a simple procedure for systematically ferreting out such information.

Another technique for generating useful 'redundant' constraints is to explicitly derive the convex hull of a select (and relatively simple) subset of the set of all constraints. An illustration of this technique has been presented by Geoffrion and McBride [16].

Additional constraints discovered by one or another of the means described above can be used in a number of other ways beyond simply appending them

to the LP relaxation for bounding; for instance, they can be used to quickly eliminate variables at various points in the search tree, and to strengthen so-called penalties. In addition to [42], confirmation of their value for large practical applications can be found in Refs. [16] and [13, see especially Section 5], and in Ref. [4].

Thus, the integer programming modeler must learn that economizing on the number of constraints in the representation of a model can be a sin rather than a virtue. Economizing on the number of integer variables, however, is usually very desirable. New tricks are found from time to time which permit more compact transformations of combinatorial problems or nonlinear integer programs into linear integer programs, usually with the help of binary variables. For instance, an improved technique for linearizing quadratic forms in n binary variables by introducing only n continuous variables and $4n$ constraints is given in [17]. Having said this, a word of caution must be added that it is often very inefficient to transform an inherently combinatorial or nonlinear problem into a linear integer program. One should also keep in mind that there are conspicuous exceptions to the rule that the number of integer variables should be minimized. Sommer [37] argues convincingly (see Cases 1 and 4) that important decision alternatives should always be represented explicitly by individual integer variables rather than implicitly in terms of several variables, even if this means introducing some additional binary variables. The reason has to do with the fact that available IP codes can branch only on explicit integer variables (although many other kinds of branching are possible); and it is usually preferable to branch on the 'most important' decision alternatives early in the branch-and-bound search.

3. ALGORITHMS

Many intriguing algorithmic ideas continue to be proposed or developed in the literature, but relatively few are put to serious computational test. It is therefore difficult to know which developments are of truly practical value. In the near term, at least, it is probably safe to limit attention to contributions that are capable of enhancing the acknowledged general supremacy of the LP-based branch-and-bound approach.

A great deal remains to be learned about utilizing the broad tactical flexibility inherent in LP-based branch-and-bound. An important paper containing a wealth of ideas and computational evidence on penalties, priorities, branching criteria, 'pseudocosts', and related topics is by Forrest *et al.* [11]. Other substantial computational studies in a similar vein are by Mitra [28], Breu and Burdet [6] and Piper [32].

It would appear that there are two distinct kinds of evolution going on as regards the current generation of commercial software. One is a pragmatic

gravitation toward the incorporation and regular use of improved tactical options (e.g. better 'pseudocosts') for general mixed integer linear programs. This kind of gradual evolution, while an essential part of the process of assimilating a new commercial technology on a widespread basis, has probably reached a stage of diminishing returns. Further substantial increases in computational effectiveness will probably depend primarily on a different kind of evolution: innovations for the strategic exploitation of special problem structure. The outstanding example of this is the treatment of 'special ordered sets' by UMPIRE (see [11] and the references therein). The first type of special ordered set deals with multiple choice constraints over 0–1 variables, while the second deals with the structure that arises when economies-of-scale and other non-linearities are treated to piecewise-linear approximation. Both of these structures occur in a wide variety of practical problems, and have been effectively exploited so as to yield substantially improved computational efficiency. Beale and Forrest [3] have made a noteworthy contribution in this vein. It is shown how automatic interpolation can be carried out so as to avoid having to preselect the piecewise-linear approximation with the second type of special ordered set.

Another strategic development of the second evolutionary kind is Lagrangean Relaxation, which can improve the standard bounds available from linear programming (see Geoffrion [12]). It is based on a formal application of Lagrangean duality ideas and the empirical observation that the 'duality gap' tends to be very small in many applications. This technique can exploit a variety of special but common types of constraints so as to produce substantially improved bounds, penalties and cutting-planes at modest computational cost. For instance, it was found in [16] for four real facility location problems that Lagrangean Relaxation bridged an average of 60 per cent of the gap between the LP optimal value and the integer optimal value for a nominal computational cost, with a maximum average potential of 97 per cent if greater effort is expended. Moreover, the penalties associated with fractional variables were one or two orders of magnitude better than those available from standard formulae yet were less expensive to compute. Of course such favorable results may not be obtained for other applications, but independent results by other investigators for other problems are encouraging. Lagrangean Relaxation has been used successfully by M Held and RM Karp in their well-known and successful attack on the traveling salesman problem. Another nice illustration is provided by Ross and Soland [36]. They have succeeded in solving highly structured problems with several thousand binary variables in a few seconds on a large machine. Still another successful application is reported by Fisher [9]. See also the ambitious paper by Fisher *et al.*[10]. It is interesting to note that these other investigators have all used Lagrangean Relaxation *alone*, rather than as an adjunct to linear programming. It can be used either way.

A point made earlier in these notes was that improved bounds are frequently possible through improved model representation practices. My impression is

that there is a lot of room for improvement here. Some of the tricks mentioned in that context could even be automated as part of the preprocessing of a problem prior to its solution. To the extent that this leads to the introduction of additional constraints, it reminds one of the not infrequently mentioned suggestion to incorporate cutting-planes into branch-and-bound as a means of improving the tightness of the usual LP relaxation, and to facilitate effective branching (cf. Section IV of [15]). Many varieties of cutting-planes have been proposed over the years, the vast majority of them untested, but excellent results have been obtained using just the original Gomory mixed integer cuts of 1960 in a hybrid branch-and-bound/cutting-plane implementation for solving the master problems in [13]. Very few studies of this type of hybrid algorithm have been reported. One modest study by Knuts [24] has appeared on a hybrid version of Gomory cutting-planes with Balas' algorithm. The hybrid algorithm proved superior to the pure algorithms on which it was based.

More than a decade ago, JF Benders proposed an approach for iteratively decomposing mixed integer problems into their continuous and integer parts, each of which can then be solved by appropriate specialized algorithms. Progress has been made recently in finding variants of this approach which work well for specific classes of problems. A large scale implementation and its successful application to a practical problem is described by Geoffrion and Graves [13]. An interesting point discussed in [13] is that the choice of the mathematical representation of a model can have a decisive influence on computational efficiency; moreover, the most powerful representation for this class of problem is too unwieldy to be attacked directly via branch-and-bound. Other promising variants of Benders' approach are reported in [8] and [27]. In addition, a successful practical largescale application using Benders' decomposition for the LP relaxations within branch-and-bound is described by Austin and Hogan [2].

IV. *Post-optimal and reoptimization capability*

It is surprising that so little attention has been devoted to post-optimal analysis and reoptimization capabilities for integer programming. These topics, so well developed and widely used in the domain of linear programming, are needed even more acutely for practical applications of IP models. This is because the intrinsic nature of integer programming generally calls for greater caution in interpreting the results and generally requires more runs in order to do justice to a real application, especially strategic applications. For instance, [13] identifies and discusses eight types of runs necessary to properly carry out a facility location study. Most of these actually call for an entire sequence of related runs, each of which must be solved to a very high degree of optimality in order to be meaningful (since it is the *differences* between the solutions of the runs in a series that matters). Thus there is a clear need for true reoptimization and post-optimal capability.

The first attempt to build such capability into a branch-and-bound algorithm

was carried out by Roodman [34] in the context of Balas' additive algorithm. The basic concept in [34] is that for the purposes of postoptimization it is sufficient to limit consideration to the partial solutions that were fathomed in the course of the initial solution of the problem. Similar developments in the more pertinent context of LP-based branch-and-bound are given in [35]. A relatively comprehensive study of this budding field has just been completed by Nauss [30]. It includes extensive computational experience with several classes of problems, some of it of independent interest as further experimental confirmation of the value of the Lagrangean relaxation ideas mentioned earlier.

Benders' decomposition method presents a different kind of opportunity for reoptimization: the so-called cuts generated to solve one problem can be saved and modified at little cost so as to be useable in many modified versions of the same problem. This point is discussed in detail in Section 2.4 of [13] and has also been recognized and employed successfully (60–80 per cent reduction in computing time) by Noonan [31]. With a little care, a number of cutting-plane algorithms also lend themselves to reoptimization, although no computational experience has yet been reported.

ACKNOWLEDGEMENTS

Partially supported by the National Science Foundation under Grant GP-36090X and by the Office of Naval Research under Contract N00014-69-A-0200-4042. The author would like to express his appreciation to David C Sommer and William W White for their valuable comments on an earlier version of this note. Reproduction in whole or in part is permitted for any purpose of the United States Government.

REFERENCES

1. Armstrong RD and Sinha P (1974) Application of quasi-integer programming to the solution of menu planning problems with variable portion size. *Mgmt Sci.* **21**(4).
2. Austin LM and Hogan WW Optimizing procurement of aviation fuels. *Mgmt Sci.* forthcoming.
3. Beale EML and Forrest JJ (1973) Global optimization using special ordered sets. Scientific Control Systems Ltd.
4. Beale EML and Tomlin JA (1972) An integer programming approach to a class of combinatorial problems. *Math. Progm.* **3**(3), 339–344.
5. Bradley GH, Brown GG and Graves GW (1975) A comparison of storage structures for primal network codes. Paper presented at the ORSA/TIMS National Meeting, Chicago, May.
6. Breu R and Burdet C (1974) Branch and bound experiments in zero-one programming. *Math. Progm. Study* 2.
7. Computer Research Center for Economics and Management Science (1974) SESAME: design and capabilities overview. Document D0075, National Bureau of Economic Research, Inc., Cambridge, Mass. March. (Public domain)

8. ETSCHMAIER MM and RICHARDSON RJ (1973) Improving the efficiency of Benders' decomposition algorithm for a special problem in transportation. Tech. Rep. No. 14, Dep. of Industrial Engineering, University of Pittsburgh, May.
9. FISHER ML (1974) A dual algorithm for the one-machine scheduling problem. Tech. Rep. No. 243, Dep. of Operations Research, Cornell University, December.
0. FISHER ML, NORTHUP WD and SHAPIRO JF (1974) Using duality to solve discrete optimization problems: theory and computational experience. Rep. 7409, Graduate School of Business, University of Chicago, February. *Math. Progm.* forthcoming.
11. FORREST JJH, HIRST JPH and TOMLIN JA (1974) Practical solution of large mixed integer programming problems with UMPIRE. *Mgmt Sci.* **20**(5), 736–773.
12. GEOFFRION, AM (1974) Lagrangean relaxation for integer programming. *Math. Progm. Study* 2, 82–114.
13. GEOFFRION AM and GRAVES GW (1974) Multicommodity distribution system design by Benders decomposition. *Mgmt Sci.* **20** (5), 822–844.
14. GEOFFRION AM and GRAVES GW (1975) Scheduling parallel production lines with changeover costs: practical application of a quadratic assignment/LP approach. Working Paper No. 231, Western Management Science Institute, UCLA, April.
15. GEOFFRION AM and MARSTEN RE (1972) Integer programming algorithms: a framework and state-of-the-art survey. *Mgmt Sci.* **18**(7), 465–491.
16. GEOFFRION AM and MCBRIDE RD (1972) The capacitated facility location problem with additional constraints. Paper presented at the Joint National Meeting of AIIE, ORSA and TIMS, Atlantic City, N.J. November.
17. GLOVER F (1974) Improved linear integer programming formulations of nonlinear integer programs. *Mgmt Sci. Rep.* No. 72–8, School of Business, University of Colorado, Rev. April. *Mgmt Sci.* forthcoming.
18. GLOVER F, KARNEY D, KLINGMAN D and NAPIER A (1974) A computation study on start procedures, basis change criteria, and solution algorithms for transportation problems. *Mgmt Sci.* **20**(5), 793–813.
19. GONCHARSKY RS, RAUCH A and WHITE WW (1973) Large scale mathematical programming in an APL environment. Tech. Rep. No. 320–3027, IBM Philadelphia Scientific Center, October. (Program is not available outside of IBM.)
20. GRAVES GW and MCBRIDE RD (1973) The use of inherent triangularity in network problems. Paper presented at the 44th National Meeting of ORSA, San Diego, Calif. November.
21. GRAVES GW and MCBRIDE RD (1975) The factorization approach to large-scale linear programming. Working Paper No. 208, Western Management Science Institute, UCLA, Rev. May. *Math. Progm.* forthcoming.
22. HAMMER PL and NGUYEN S (1972) APOSS: a partial order in the solution space of bivalent programs. C.R.M.-163, Centre de Recherches Mathematiques, Universite de Montreal, April.
23. KARNEY D and KLINGMAN D (1973) Implementation and computational study of an incore out-of-core primal network code. Rep. C.S. 158, Center for Cybernetic Studies, University of Texas, Austin, October. *Ops Res.* forthcoming.
24. KNUTS L (1973) Combining two integer programming algorithms. *Acta Acad. abo.* Ser. B, **33**(12), 1–8.
25. MANAGEMENT SCIENCE SYSTEMS (1974) MPS III/OL user manual. Rockville, Maryland. (Proprietary system)
26. MCBRIDE RD (1973) Factorization in large-scale linear programming. PhD Dissertation University of California, Los Angeles, June.
27. MCDANIEL D and DEVINE M (1974) Alternative relaxation schemes for Benders' partitioning approach to mixed integer programming. School of Industrial Engineering, University of Oklahoma, April.
28. MITRA G (1973) Investigation of some branch-and-bound strategies for the solution of mixed integer linear programs. *Math. Progm.* **4**(2), 155–170.
29. MULVEY JM (1975) Developing and testing a truly large-scale network optimization code. Paper presented at the ORSA/TIMS National Meeting, Chicago, May.

30. NAUSS RM (1974) Parametric integer programming. PhD Dissertation, Graduate School of Management, University of California, Los Angeles, December. Available as Working Paper No. 226, Western Management Science Institute, UCLA.
31. NOONAN F (1973) Optimal investment planning for electric power generation. PhD Dissertation, Dep. of Industrial Engineering and Operations Research, University of Massachusetts, September.
32. PIPER CJ (1974) Implicit enumeration: a computational study. Working Paper No. 115, School of Business Administration, University of Western Ontario, London, Canada, May.
33. RADKE M (1975) Sensitivity Analysis in Discrete Optimization. PhD Dissertation, Graduate School of Management, UCLA, September.
34. ROODMAN GM (1972) Postoptimality analysis in zero-one programming by implicit enumeration. *Nav. Res. Logist. Q.* **19**(3), 435–447.
35. ROODMAN GM (1973) Postoptimality analysis in integer programming by implicit enumeration: the mixed integer case. Working Paper, Amos Tuck School of Business Administration, Dartmouth College, October.
36. ROSS GT and SOLAND RM (1975) A branch and bound algorithm for the generalized assignment problem. *Math. Progm.* **8**(1).
37. SOMMER D (1972) Computational experience with the Ophelie mixed integer code. Control Data Corp., Minneapolis, Minn. May.
38. THOMAS GS, JENNINGS JC and ABBOTT P (1973) A blending problem using integer programming on-line. Paper presented at the *VIII International Symposium on Mathematical Programming, Stanford University, August* (authors at Scicon, Ltd. London).
39. TOMLIN JA (1972) Survey of computational methods for solving large scale systems. Tech. Rep. 72–25, Dep. of Operations Research, Stanford University, October; reprinted in *Proceedings of the IEEE Conference on Decision and Control and 11th Symposium on Adaptive Processes, December.*
40. WHITE WW (1973) A status report on computing algorithms for mathematical programming. *Computg. Surv.* **5**(3), 135–166.
41. WILLIAMS AC (1973) Some modeling principles for M.I.P.'s. Paper presented at the *VIII International Symposium on Mathematical Programming, Stanford University, August* (author at Mobil Oil Corp. Princeton, N.J.).
42. WILLIAMS HP (1974) Experiments in the formulation of integer programming problems. *Math. Progm. Study* 2, December.
43. WINKLER C (1974) Basis factorization for block-angular linear programs: unified theory of partitioning and decomposition using the Simplex method. Tech. Rep. SOL 74–19, Systems Optimization Laboratory, Stanford University, December.

ADDRESS FOR CORRESPONDENCE: *Professor A Geoffrion, Western Mgmt Science Institute, University of California, Los Angeles, Calif 90024, USA.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

The 'Optimum' Number of Beds in a Coronary Care Unit

A A SISSOURAS

B MOORES

(*Received March 1975; in revised form August 1975*

A study of the Coronary Care Services provided in the community of Salford has been underway now for more than 3 years. The effects of the introduction of a mobile cardiac ambulance on the community and the hospital are being investigated. This paper reports on one important issue stemming from the present study namely how many beds a CCU ought to develop according to medical and operational criteria. On the one hand an excess of beds results in under-utilisation whilst on the other a lack of beds results in the treatment of some patients in the CCU being truncated.

INTRODUCTION

CORONARY Care Units (CCU) engage resources, particularly human, which are not only expensive but scarce. A typical six-bedded unit will involve a capital expenditure of around £30,000 and annual running costs of the order of £50,000. Admissions to the hospital will undoubtedly increase due to the existence of a CCU and the rate of referral by general practitioners has been found to increase further when a Mobile Cardiac Ambulance is attached to the system. These effects have been experienced at Hope Hospital, Salford, where such an upsurge of admissions has affected the capacity and operational characteristics of the hospital. The results of the following simulation exercise provides a guide to the planners of prospective CCU's as to what number of beds is most appropriate for various projected admission rates.

The system of coronary care services and its operational characteristics

The system to be analysed is centred on the Hope Hospital Coronary Care Unit. Since the introduction of the mobile unit 55 per cent of patients are

admitted via this channel and the arrival pattern from *all* sources is shown in Fig. 1, together with the expected distribution based on the assumption of an underlying Poisson process.[1]

The patient is typically scheduled to remain in the unit for three days but there are, naturally, deviations from this nominal length of stay. These deviations might reflect the severity of the disease or the bed availability in either the CCU or the wards. When the CCU is fully occupied any new arrival results in a patient currently in the unit having to be transferred as obviously no queues can be permitted at that stage. Likewise, when the unit is not fully occupied, a patient due to be transferred may stay on when no bed is available in the wards. The actual length of stay distributions used in the simulation are the empirical ones exhibited in Fig. 2 based on 330 patients over a ten month period.[2] Separate distributions for myocardial infarctions and non-myocardial infarctions were used. Given the obvious differences, it is clearly necessary to discriminate between the two as their relative incidence can change quite dramatically upon the introduction of new facilities.[3]

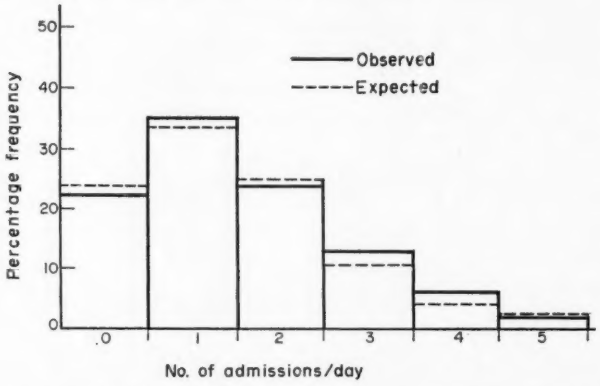


FIG. 1. *Observed and expected distribution of the number of admissions to CCU per day.*

[1] The arrival patterns for the two sources: Mobile Unit and Casualty, which together now account for more than three-quarters of the admitted patients, have been examined independently and have both been found to conform to a Poisson process.

[2] At the time the CCU had recently been extended from three to six beds but the increased influx resulting from the introduction of the mobile cardiac ambulance had not yet materialised. The consequent low utilisation of the unit implies that the frequency with which patients had to be discharged due to pressure on beds was very low but there would be some bias at the other end of the scale due to the stay in the unit being artificially extended. To secure a 'natural' length of stay distribution would require the consultant to indicate at the time of discharge from the unit whether the discharge is natural, forced or stretched artificially. If forced or stretched then he would be asked to estimate what would constitute the natural figure.

[3] At Salford the proportion of patients with proven myocardial infarctions decreased only slightly upon the introduction of the mobile cardiac ambulance but in Belfast, where the mobile unit answers calls not only from general practitioners but also from other sources, the proportion of false positives has been as high as 50 per cent.

Vol. 4
1976

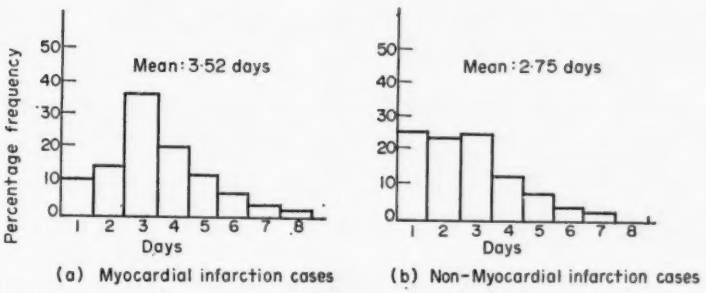


FIG. 2. *Distributions of length of stay in the CCU*

Simulation procedure

Each time a new arrival is generated, it is randomly designated as a 'myocardial' or 'non-myocardial' infarction case, in proportion to the relative incidence (60/40) and a length of stay allocated from the appropriate distribution (Fig. 2). Whenever an arrival coincides with the unit being full a patient is ejected. The patient ejected is the one with the least time remaining prior to his natural discharge. Thus, if a patient, originally 'designated' as a 5-day stay, had been in the unit for 3 days, whilst another designated three had been in for two, it would be the latter who would be ejected by virtue of the fact that he had only 1 day of his projected stay remaining in contrast to the other's two. A sub-routine generates deaths according to the survival function prevailing in the unit which is shown in Fig. 3.

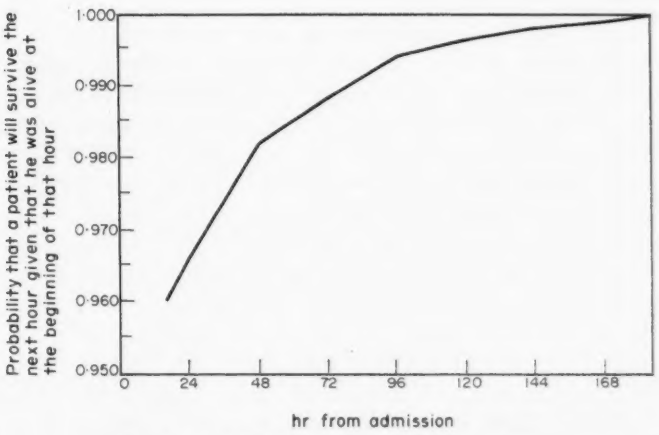

Probability that a patient will survive the next hour given that he was alive at the beginning of that hour
hr from admission

FIG. 3. *Survival function of patients in the CCU*

The operational policies, the parameters and the probability distributions considered are as follows:

(1) Admit every patient 'arriving' in the unit thus allowing for no queues. The arrivals are Poisson with the average daily rate taking on values of 0·25, 0·50, 0·75, ..., 4·00.

(2) Distinguish MI and non-MI cases in the ratio 60:40.

(3) If, upon arrival, all beds are occupied, the patient with the shortest scheduled time left in the unit is ejected.

(4) Assign to each patient a projected length of stay in the unit. The length of stay distributions utilised are different for MI and non-MI cases and are those shown in Fig. 2.

(5) Apply a 'die' and 'live' process to the patients in the unit according to the survival function depicted in Fig. 3.

(6) Transfer patients out of the CCU whenever their pre-assigned length of stay comes to an end.

(7) The number of beds in the unit can be varied from 1 to 10.

Vol. 4
1976

RESULTS

The primary purpose of the study was to investigate the level of 'ejections' in a unit operating under different rates of arrivals and varying number of beds. Such information coupled with that on 'bed utilisation' can be incorporated into planning decisions concerning the most appropriate number of beds in the CCU. The behaviour of the whole coronary system was simulated for 350 days for each combination of bed and arrival rate. The data from the first 150 days was not used in the eventual analysis but allowed the system to stabilise. A typical print-out is shown as Table 1.

The measures which were extracted from the results of the simulation are as follows:

(i) *The 'ejection' level*—the proportion of all arrivals who are eventually ejected from the CCU. This is, of course, equivalent to the likelihood that a new patient will find the unit full. Figure 4 depicts the relationship between this ejection level and the rate of arrival.

TABLE 1. A TYPICAL DEVELOPMENTAL COMPUTER PRINT OUT OF THE RESULTS OF A 100-DAY RUN

NO. OF BEDS = 6 RATE OF ADMISSION = 1·5

HOUR AT WHICH DEATH OCCUR SINCE ADMISSION OF PATIENT
1: 20HR 2: 30HR 3: 21HR 4: 57HR 5: 44HR 6: 52HR 7: 35HR
8: 9HR 9: 9HR 10: 2HR 11: 4HR
TOTAL NO. OF DEATHS = 11

FREQUENCY DISTRIBUTIONS

I	ARRIVALS	TRANSFERS OUT	BEDS OCCUP.	EJECTS	'CUT CARE'
0	23	37	0	75	0
1	32	39	4	13	34
2	21	15	13	5	14
3	17	7	11	4	0
4	4	2	18	2	0
5	3	0	15	1	0
6	0	0	39	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
SUM:	156	98	444	48	62
MEAN:	1·53	0·98	4·44	0·48	0·62

BED UTILISATION = 0·74

(ii) *The 'cut care'* per ejected patient expressed as the average number of days the 'normal' stay of the patient was truncated. (Values of 'cut care' are seen in Fig. 5.)

(iii) *The average bed utilisation* (Numerical values of bed utilisation are seen in Fig. 6.)

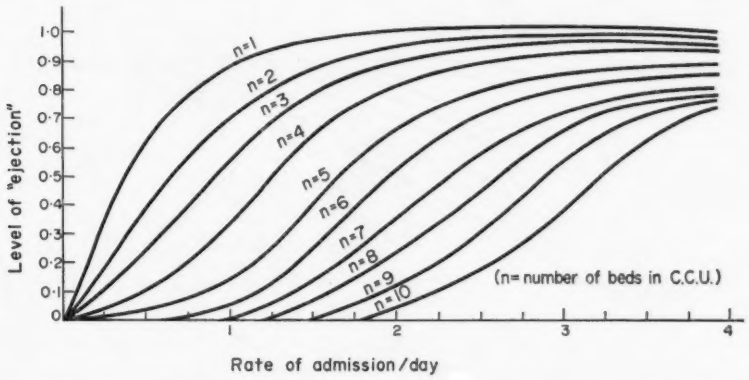


FIG. 4. *Proportion of patients ejected from the CCU for varying number of beds and varying admission rates.*

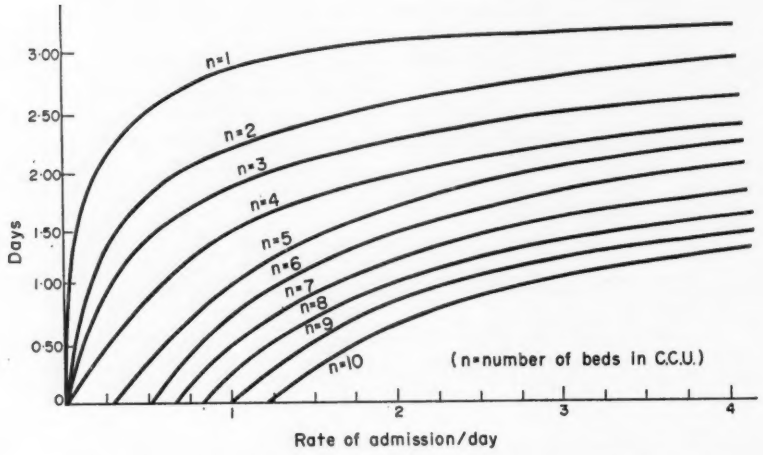


FIG. 5. *Average number of truncated days per ejected patient ('cut care').*

From these figures it can be seen, for example, that when there are five beds in the unit and arrivals average two per day, 67 per cent of all patients have to be ejected and on average this group have their stay reduced by 1·71 days. The associated average occupancy of the unit is 87 per cent.

The above measures provide a set of suitable indices for decisions relating to the operation of a CCU. If admissions could be controlled by the simple expedient of refusing admission when the unit is fully occupied then it is quite easy to reach a high level of bed utilisation and no truncation of treatment. In the real world, however, patients are only rarely, if ever, turned away from the unit. This high level of acceptance can only be achieved by what might at first sight appear to be an over-provision of beds. The bed utilisation factor would,

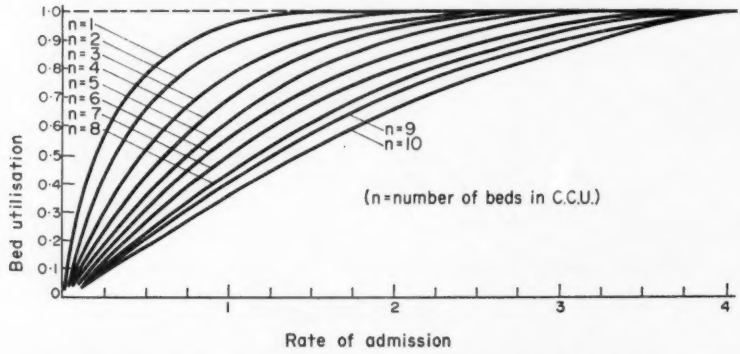


FIG. 6. *Bed utilisation for varying number of beds and different admission rates.*

in that case, become lower and ways would have to be found for balancing the acceptance level and the bed utilisation. By incorporating the 'ejection' level, a better compromise is perhaps possible. Although a number of patients are 'pushed' out of the unit before the completion of their specified treatment, this is likely to be seen as a more acceptable policy than refusing entry. The eventual decision has to represent a compromise between resource utilisation and level of service but the information contained in Figs. 4–6 enables the full implication of any decision to be pre-determined.

From Fig. 4 it can be seen that if the average rate of admission is two per day, then at least 9 beds are required if the 'ejection' level is not to exceed 20 per cent. Likewise for a daily admission rate of three, the number of beds required to achieve the same level of service must exceed ten. To secure an ejection level as low as 10 per cent with a rate of admission higher than one per day, at least 6 beds are required.

The six-bedded Salford Unit, for example, which accommodated 1·45 admissions per day in 1972, is seen to exhibit an 'ejection' level of approximately 23 per cent. The average 'cut care' per ejected patient is 1·17 days whilst the bed utilisation is approximately 70 per cent. For this rate of admissions an 'ejection' level lower than 20 per cent can only be secured by the provision of at least 7 beds. The 'cut care' value would then be around one day and the bed utilisation 65 per cent.

CONCLUSION

This paper illustrates how the tool of simulation can be brought to bear on the task of determining an appropriate number of beds to be provided in a coronary care service which can either be an autonomous coronary care unit or an integrated system comprising such a unit together with a ward into which the CCU patients are discharged. Clearly, in any other setting the operating parameters will differ but the methodology would be identical. This particular problem is, of course, but one illustration of the perennial one of reconciling the levels of service and resource utilisation.

ADDRESS FOR CORRESPONDENCE: *Dr B Moores, Department of Management Sciences, UMIST, POB 88, Manchester M60 1QD, UK.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Simulation of R&D Investment Strategies[1]

A DOUGLAS BENDER

EDMUND B PYLE III

WILFRED J WESTLAKE

BRYCE DOUGLAS

(*Received March* 1975; *in revised form May* 1975)

Simulation has been applied to a descriptive model of the R&D process and its relationship to other components of a commercial system with the objective to: (1) consider the size of the R&D expenditure in terms of future profit growth; (2) estimate the rate of return on the R&D investment; (3) compare the benefits of investment strategies designed to improve profit growth and return on investment; and (4) assess the consequences of anticipated extra corporate environmental futures on the attractiveness of the investment in R&D. With the use of this econometric model a number of major issues may be addressed in terms of their effect on R&D profitability and return on investment. These include: (1) the size of investment in R&D; (2) the level of research productivity; (3) strategies for development of new products worldwide; (4) rising R&D costs and inflation; (5) the probability of technical success; (6) market potential; (7) profit margin; and (8) length of product exclusivity.

ONE OF the most important decisions the management of a research-intensive corporation must make is how much it should invest in R&D (research and development) activity. Traditionally, the size of the R&D budget in a given year is expressed as a percentage of current sales level. For example, in the paper and machinery industries in the US R&D budgets represent approximately 1 per cent of sales; in the pharmaceutical industry this figure may exceed 10 per cent. Ideally, decisions regarding the level of expenditure on R&D should be

[1] A preliminary report was presented at the 7th Annual Winter Simulation Conference, Washington, D.C., January 1974 and appeared in *Proceedings of the 7th Winter Simulation Conference*, Vol. 1, p. 121, 1974.

related to expected future sales and the profits to be derived from the results of the R&D effort and with knowledge of the level of risk associated with a given level of investment [1].

In the past several years, simulation has emerged as a powerful analytical tool [2] with application to a variety of economic, financial, manufacturing, and R&D models [3–7]. The present report describes the application of simulation to a descriptive model of the R&D process and its relationship to other components of the commercial enterprise. The model was designed as one approach to the consideration of:

the size of the R&D expenditure in terms of future profit growth;

the rate of return on the R&D investment as a basis for comparing investment alternatives;

the comparative benefits of strategies designed to improve profit growth and return on investment, i.e. productivity, cost, product development strategies, etc.;

the consequences of anticipated extra-corporate environmental futures on the attractiveness of investment in R&D (i.e. inflation, regulations, duration of product exclusivity and profit margins).

Vol. 4
1976

In addition to a description of the model and methodology, a number of studies are described which are made possible by the availability of such an econometric system and the ability to assess risk through the application of simulation.

THE MODEL

General orientation

The research and development process in the ethical pharmaceutical industry incorporates many varied activities. These may be grouped into the following basic elements: the synthesis or isolation of novel chemical entities, evaluation of these compounds in biological test systems designed to discover desirable biological activities, assessment of the safety and efficacy in man of compounds judged to merit clinical trial, and finally securing permission to market the new drug. This process of discovery, development and marketing is expressed schematically in Fig. 1. In addition to projects arising from research, projects enter the development phase from other sources such as direct acquisition or joint ventures. Once a development project is established, a development/ marketing strategy is selected based on the predicted product profile,

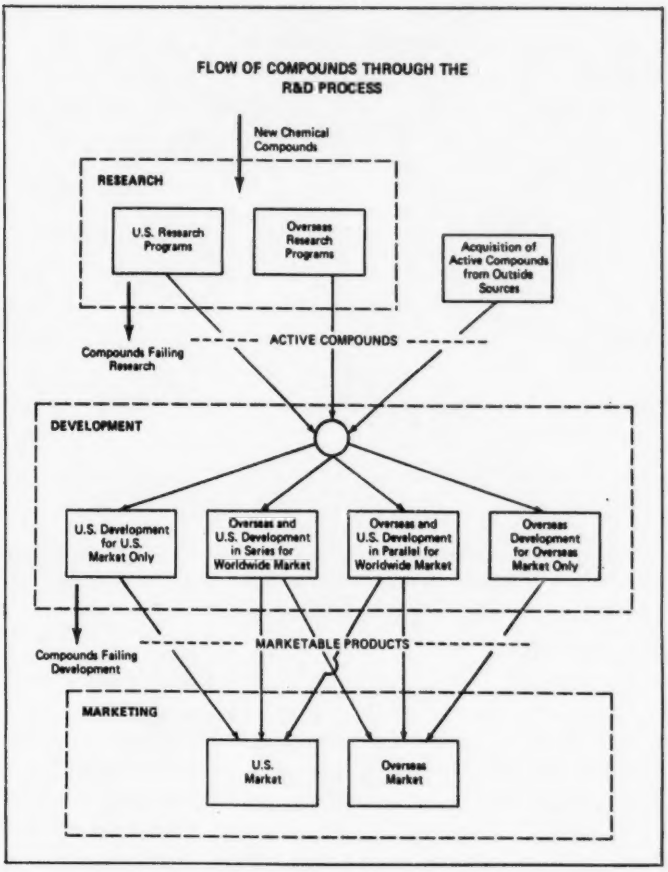


FIG. 1. *Pharmaceutical R&D process.*

anticipated market, technical risk, and available resources. Of all the compounds entering the development process, a small fraction survive and become marketable products.

The model is made up of *investment* and *profit* components. The *investment* component describes the extent and nature of the research and development commitment in terms of dollar spending, productivity, and probability of success. The three modules of this component are research, clinical evaluation, and development. Each is described in terms of success rates, duration and level of spending. From these data it is possible to calculate the rate at which new projects are produced and then to assign costs at each stage in the process. The *profit* component is a description of the performance of the product in the market place. New products resulting from the investment in R&D enter the

market under varying conditions of uncertainty. To accommodate this uncertainty, distribution curves are required for sales and pretax profit margin. From the expenditure component and the profit component a net cash flow stream may be generated (Fig. 2) which yields the profit contribution in any given year and makes possible the determination of the internal rate of return and net present value.

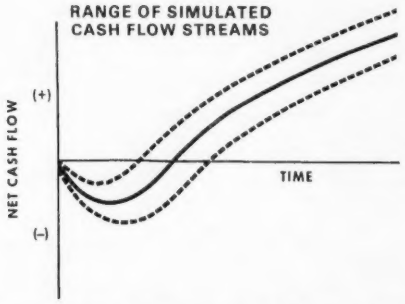


FIG. 2. *Schematic representation of the mean and quartile range of cash flow stream generated by the model using a finite period for the R&D investment.*

Mathematical considerations

Due to the complexity of the entire research and development process and the uncertainties associated with future events, it was apparent at the outset that a closed-form mathematical representation of the system was not possible. In addition, representing the system through a series of empirically derived relationships was not considered feasible due to the stochastic nature of many of the system elements. Since the behavior of many of the elements are known in terms of historically derived probability distributions, and since the path through the R&D process for a particular chemical compound can be treated as a series of discrete events, a simulation model was considered to be the logical choice to represent the system. Conceptually, the model employs Monte-Carlo techniques to determine when each compound is ready to begin to enter the development process, which development/marketing strategy will be employed, whether or not each compound survives development, and the sales and profit margin achieved for a marketable product. A flow chart of the model is shown in Fig. 3.

The queue of random compound arrivals is generated by considering each source. Actually, a set of random arrivals is generated for each source independently and then merged to form the queue of compounds for input to the development process. This is accomplished by assuming the flow of compounds from each source to be completely independent and random events, i.e. a Poisson process with parameter 'α' (the mean compound arrival rate). A

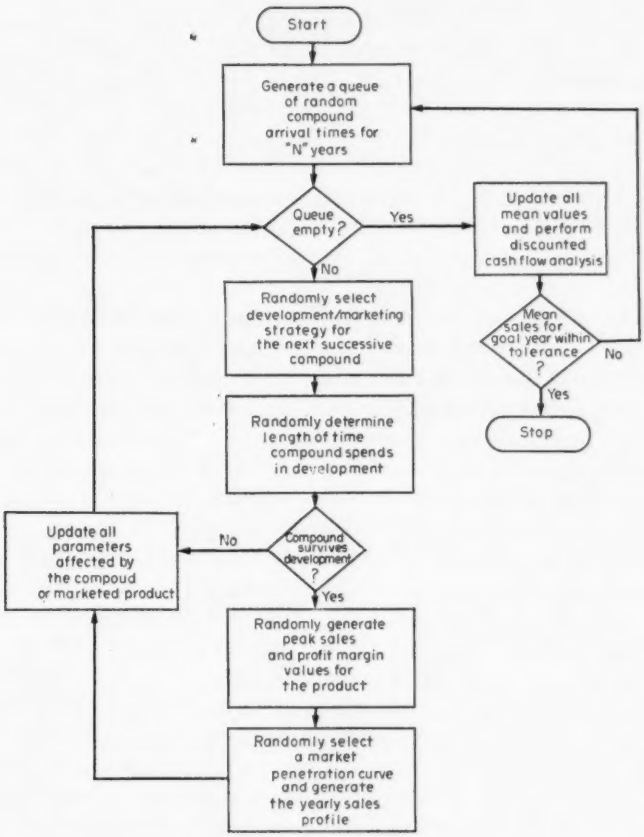

Start
Generate a queue of random compound arrival times for "N" years
Queue empty?
Yes
No
Update all mean values and perform discounted cash flow analysis
Mean sales for goal year within tolerance ?
No
Yes
Stop
Randomly select development/marketing strategy for the next successive compound
Randomly determine length of time compound spends in development
Compound survives development ?
No
Update all parameters affected by the compoud or marketed product
Yes
Randomly generate peak sales and profit margin values for the product
Randomly select a market penetration curve and generate the yearly sales profile

FIG. 3. *System flow chart.*

separate stream of arrival times is generated for each source. In the case of an acquisition, 'α' is an input variable. In the case of original corporate research, 'α' is determined from the research spending, and productivity, all of which are input variables.

As the research spending increases, the number of new projects in the input queue also increases. The length of the input queue in years and the first year of the 25 year span of interest are supplied as input. In this manner, the value of an 'n' year research spending over a 25-year span may be evaluated.

Once the input queue of potential development projects is generated, the model begins selecting them for clinical evaluation followed by full scale development, one at a time, until the queue is exhausted. This entire process assumes that a project is started in development as soon as it is available. For each project selected, the model randomly picks a development/marketing

strategy. The selection of a particular strategy is made based on the occurrence fraction specified for each. The occurrence fractions are supplied as input and reflect the relative amount that each strategy is to be utilized during that particular simulation. They can have any value from 0 to 1 for each individual strategy but the sum of all occurrence fractions must be equal to 1. We have identified seven basic development/marketing strategies which vary in the degree to which activities are planned in series and/or in parallel. Each is associated with a different number of years for development (N), yearly cost allocation over the development period and probability (Ps) that a new product emerges from the process.

After selecting a development marketing strategy, the model determines how long the project survives the development process. It does this on a probabilistic basis by assuming that the first year of development is completed with certainty, and the probability of completing each successive year thereafter decreases linearly with time until 'P_s' is reached for year 'N'. After randomly determining the number of years of survival of the project in the development process the model adds the development costs for this number of years into the appropriate yearly totals and proceeds to pick up the next project in the queue.

If the project survives the 'N' years, it is assumed to successfully complete development and becomes a marketable product. The model next randomly determines a peak year sales figure for sales for a given market or individual market segments by generating a random number from a Weibull distribution whose upper and lower confidence limits are determined from historical data. The Weibull distribution was selected over the normal distribution because it begins at zero and, therefore, negative values do not pose a problem. Once the peak sales figure is obtained, the model randomly selects a market penetration rate. The model has three penetration rates: slow, normal, and fast. These correspond to products which face stiff competition, average competition and little or no competition respectively. Again, a random process based on these percentages determines which penetration rate is used. Associated with each penetration rate is the curve of percent of peak sales as a function of time. Hence, once the penetration rate has been determined the yearly sales figures are obtained by applying the appropriate market penetration curve to the peak sales. The model next randomly determines a pretax profit margin value generating a random number from a Weibull distribution whose upper and lower confidence limits were determined from historical data. The profit margin value is expressed as a percentage of sales and, hence, must be applied to the various sales curves in order to obtain absolute values. Finally, the model inflates the yearly sales, profit margins and development costs by applying factors which are supplied as input and adds the resulting values into the appropriate yearly totals. It then proceeds to consider the next project in the queue.

When all projects in the queue have been exhausted, the model updates the mean yearly values of development costs, sales and profit margin. In addition

Vol. 4
1976

to these parameters, the model also updates mean yearly values for the number of projects starting development, failing development and being marketed for each year in the 25-year span of interest. Next, a net cash flow stream for the 25-year span is generated by subtracting the yearly research and development costs from the pretax profit. A discounted cash flow analysis is performed on this cash flow stream and the Net Present Value (NPV) and Internal Rate of Return (IRR) obtained. Following this procedure, the model reinitiates all variables and proceeds to generate a new queue of development compound arrivals for another run. This process is continued until some minimum number of runs have been completed, after which the model computes how many runs are necessary to achieve some degree of stability in the mean sales values. This is accomplished by utilizing a precision specification on the mean sales for some particular year of interest within the chosen 25-year span. This target year usually corresponds to some year for which a sales level goal has been established. Both the target year and the precision specification, with the precision specification being expressed as a fraction of the mean sales, are supplied as input.

ANALYTICAL STUDIES

The availability of an econometric model makes possible a number of studies with a minimum of time and effort and permits decision makers to rapidly assess the effects of a number of alternative assumptions.

The initial stimulus for the development of this model was a question dealing with the degree of consistency between the level of investment in R&D with the longer term goals of the total commercial enterprise. This question is an obvious one for a research-intensive industry when a projection such as that shown in Fig. 4 is generated. This projection shows a substantial gap between the desired future sales goal and the current product sales forecast. In addition, two other questions are relevant; namely, what is the risk associated with the investment and does the investment itself generate an adequate return.

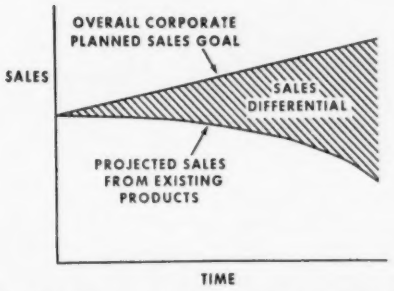


FIG. 4. *Relationship between projected goals and current product sales forecast.*

To address the size issue, the relationship between R&D expense and expected new product sales (or profits) can be developed within the framework of a given set of environmental conditions and assumptions regarding R&D productivity and probability of technical success. A relationship between R&D expenditure and new product sales in a target year along with the quartile range is shown in Fig. 5. From these data it is possible to assign a level of probability to the achievement of a given sales level for any target year and for each level of research expenditure. For example, there is a 50 per cent probability that the sales goal (line X) will be achieved or exceeded at a level of research investment designated as point A. A higher level of risk may be accepted by allocating fewer dollars to R&D: at point B, for example, there is only a 25 per cent probability that the goal will be achieved. If an organization wished to minimize its risk they could invest more (point C) in R&D so as to increase the level of probability (in the example shown, to 75 per cent) that the desired level of new product sales will occur.

The study described represents but one potential use of the model. Many additional applications are possible but they generally fall into two categories:

the development of investment strategies and management objectives;

evaluation of the potential impact of changes in uncontrollable factors in the extra corporate environment.

Vol. 4
1976

In our own studies we began by developing a control solution based on current planning assumptions and then varying each of the input parameters while

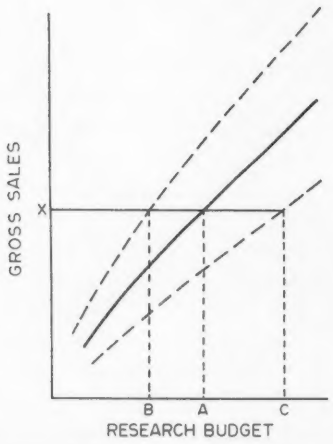


FIG. 5. *Relationship between R&D expenditure and sales generated in a future target year.*

holding the others constant, so as to determine the relative impact of each input factor on R&D profitability. In further studies the input factors were varied in different combinations in an effort to identify strategies which would achieve a more optimal investment outcome. The following are representative of the types of studies that are possible with this model.

Research productivity

By varying the level of research productivity at a given cost per productive unit it is possible to define the benefits of programs which have as their objective to specifically promote and stimulate innovation and creativity. Since scientific productivity is an issue which is always under consideration, the leverage provided by an improved level of productivity can be placed in perspective with regard to other strategies or management objectives under consideration.

Licensing and joint agreements

The model accommodates the introduction into development of potential new products acquired through licensing or which result from joint ventures. Since products from these sources differ from internally generated products in terms of development resources, costs, monetary and/or technological concessions, probability of success, market share and pretax margin, it is possible to assess the relative advantage of increasing or decreasing licensing activity as opposed to an increase or decrease in the funding of internal R&D.

Development strategies

In a complex worldwide commercial system, a number of options are open to develop a new product for each of the many worldwide markets. Each worldwide development strategy, defined in terms of the ordering of events in series and/or in parallel, will vary in cost, timing, and probability of success. Using this model it is possible to assess the relative advantages and disadvantages of any strategy under consideration. For example, it is possible to delay the development of a product for one market pending the acquisition of sufficient data elsewhere so that the probability of success is higher once the investment is committed. Thus judgment can be made regarding the trade-off between an increased probability of success and the loss of sales and profit potential that may result from a postponed market introduction or a loss in the product's competitive position that may be caused by the introduction of competitive technology during the period of postponement.

Costs and inflation

As we have noted, the model incorporates increasing costs that result either from inflation or the need for more sophisticated technology to support product discovery and development. With costs rising at significant rates around the world it is necessary to assess what portion of the anticipated growth is due to

inflation and, further, to assess the impact of inflation and the rising cost associated with more sophisticated technology on the ability to support profitable R&D at a given productivity level. In other words, at a given level of expenditure productivity will decline if the level of expenditure does not accommodate the rising cost of new technology and the costs associated with inflation.

Probability of success

The probability of successfully introducing a new product for marketing has a pronounced effect on projected R&D profitability. In other words the successful product must pay for the failure. As the attrition rate increases, the profit from a successful product must increase. Increased competition, more restrictive legislation or increased regulatory demands can significantly influence new product introductions and therefore the return on investment in R&D. Use of a model helps assess the impact of events which influence the success rate.

Sales potential, profit margin and length of product exclusivity

Studies have been conducted with the objective to determine the impact of a modification in sales potential, lower prices, and rising production costs on the return from the investment in R&D. Further, studies can be conducted which show the effect of shortening the period of product exclusivity that could come about through compulsory licensing of a product or a reduction in the duration of patent protection. These studies provide data which are useful in assessing the effects of extra corporate environmental changes on the attractiveness of investing in new technology and offer the opportunity to evaluate corporate strategies designed to offset the erosion produced by these negative influences.

Vol. 4
1976

Scenarios

Using the data generated from studies above, alternate future scenarios may be generated and their impact described in terms of pricing structure, profit margins, production costs, R&D costs, product exclusivity, probability of technical success and inflation. With these input factors varied according to the scenario description it is possible to project the impact of such an environment on the return on investment R&D to help determine what impact, if any, alternate management strategies or policies may have to improve the profitability of R&D under conditions of environmental constraints or opportunity.

RESOURCE ALLOCATION

The simulation model we have described is based on a series of discrete events. In other words, the variables in the model are treated as if they are independent of one another. For example, the level of research productivity is constant at any given level of research expenditure. Also, the ability to accept

new projects in development is independent of the arrival rate of projects from research and the ability to continually generate new sales is considered to be independent of the number of new products introduced. We have made one important assumption; namely, that these variables are independent for the range of values that have been applied in our model or that the distortion at the extremes of this variation is minimal. In dealing with allocation decisions between different avenues of investment it is necessary to make some assumptions regarding the degree of interdependence of the variables so that the allocation of resources is made in such a way as to maximize the output from the total investment. Some attempts have been made to define some of these relationships [8, 9].

CONCLUSIONS

The studies conducted using simulation models can offer corporate managements one approach to determining optimal policies and strategies for maximizing profits from investing in R&D. In a more global sense, systems and programs of this type offer the opportunity to study and assess the impact of government action, legislation, and other external factors on the relative attractiveness of this form of investment as opposed to other routes of investment.

REFERENCES

1. HAMBERT D (1966) *R&D: Essays on the Economics of Research and Development*. Random House, New York.
2. GRINYER PH (1973) Corporate financial simulation models for top management. *Omega* **1**, 465.
3. BENNETT JP (1966) Simulating alternative economic futures. In *Approaches to Long Range Forecasting* (Ed. MARTINO JP). p. 1.
4. MACASKILL JLC (1973) Computer simulation for mixed-model production lines. *Mgmt Sci.* **20**, 341.
5. FOURCANS A and HINDELANG TJ (1974) Working capital management for the multinational firm: a simulation model. *Proc. 7th Winter Simulation Conf.* **1**, 141.
6. WYNNE BE and HUTCHINSON GK (1974) Simulation of advanced manufacturing systems. *Proc. 7th Winter Simulation Conf.* **1**, 39.
7. BARKER WVH (1974) Research venture analysis. *Bull. Oper. Res. Soc. Am.* **22**, B-110.
8. GAVER DP and SRINIVASAN V (1972) Allocating resources between research and development: A macro analysis. *Mgmt Sci.* **18**, 492.
9. ANGILLEY AS (1973) Returns to scale in research in the ethical pharmaceutical industry: Some further empirical evidence. *J. Indust. Econ.* **22**, 81.

ADDRESS FOR CORRESPONDENCE: *Dr A Douglas Bender, Vice President R & D Planning Ops., Smith Kline & French Laboratories, 1500 Spring Garden Street, Philadelphia, Penn. 19101 USA.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

A Profitability Index for Alternative Research Projects

KD GLAZEBROOK

(*Received June 1975*)

Success in research is supposed to be achieved by success in just one of a number of alternative research projects. The problem is investigated on how to allocate research effort to minimise the expected cost to success. The optimal strategy is determined by a profitability index attached to each project.

1. INTRODUCTION

THE AMOUNT of money and manpower which companies have to allocate to research is limited and they seek to use the available resources in a way that will maximise their expected profitability. Much work has been done on how different research projects should be evaluated in the light of this objective [4]. In many cases, there will be inter-dependencies between available research projects, a fact which has largely been ignored in the literature. The problem considered here is one in which several research projects are available, all of which have the same objective. Success occurs and benefits accrue if any of these alternative projects is successful.

A good example of an application of such a model occurs in the area of new-product chemical research [2]. In the speculative phase of research, several different formulations may be suggested for a new product. A success will be achieved when one of these formulations has been tested successfully.

2. THE MODEL

I adopt Cranston's [1] model of a research project. A research project consists of a number of research phases which have to be done in a pre-assigned order. With each phase is associated a probability of success and a cost. A project is successful if, and only if, each of its constituent phases is successful and if this happens a reward accrues.

It is assumed that there is a single key resource, available at a uniform rate. In the above example, of new-product chemical research, the working time of a laboratory's staff of trained chemists often fills this role, though the funds available annually for the research programme could be used instead. The problem is to allocate this key resource to the individual projects in such a way that the expected cost which accrues in the time until the first project completes successfully is minimised.

Assume there are m projects. Project i has n_i phases of research, phase j being denoted by (i,j), $(1 \leqslant i \leqslant m;\ 1 \leqslant j \leqslant n_i)$. The resource cannot be allocated to $(i,j+1)$ until (i,j) has been completed. Any project may be abandoned after a research phase and may later be resumed from the point of abandonment. Such a procedure will not incur any additional cost. The probability of (i,j) being completed successfully is $P_{i,j}$ and the cost associated with it is $C_{i,j}$. If project i is successful a reward R_i accrues.

3. THE GENERAL THEORY

In the work done by Gittins and Jones [3] and Nash [4] on research project evaluation, research projects are modelled by bandit processes, so named because of their usefulness in the solution to the Multi-Armed Bandit problem. It is a fundamental result on the optimal allocation of resources within families of bandit processes which provides the solution to the problem outlined in Section 2.

Vol. 4
1976

A bandit process is a control process (Ω,U,P,C,a). Ω is the state space of the process and may be finite, countable or continuous. U is the control set, consisting of just two elements c (continue) and w (wait). P is a (probabilistic) law of motion; that is, for each $x \in \Omega$ and $u \in U$, $P(x,u)$ defines a probability distribution over some σ-field of subsets of Ω. Thus, if $X \subset \Omega$ and the state of the process at time t is $x(t)$, when control $u(t)$ is applied,

$$P[x(t+1) \in X \mid x(t), u(t)] = \int_X \mathrm{d}P[x(t), u(t)].$$

C is a cost function and a $(0 \leqslant a < 1)$ the discount factor, so that a transition at time t from $x(t)$ to $x(t+1)$ involves an incremental cost $a^t C[x(t),x(t+1)]$. The total cost accruing if the process describes a trajectory $\{x(t): t=0,1,2\ldots\}$ is

$$\sum_{t=0}^{\infty} a^t C[x(t), x(t+1)].$$

The following conditions are put on P and C.

$$P[x(t+1) = x(t) \mid u(t) = w] = 1 \quad \forall t, x(t) \in \Omega$$

$$C[x, x] = 0 \ \forall x \in \Omega \text{ when the control element is } w.$$

A family of alternative bandit processes is a control process consisting of a

set of bandit processes $\{(\Omega_i,U,P_i,C_i,a) : 1 \leqslant i \leqslant N\}$ such that at each time t $(=0,1,2\ldots)$ the control element c is applied to just one process i, w being applied to all the others. The problem studied in [3] is that of finding an optimal strategy for the control of a family of alternative bandit processes (i.e. a strategy for choosing i at each time t), in the sense of leading to minimal total expected cost.

It is shown in [3] that under certain conditions, which hold in the applications discussed here, the following theorem holds.

Theorem GJ

Given a family of alternative bandit processes $\{(\Omega_i,U,P_i,C_i,a) : 1 \leqslant i \leqslant N\}$, there exist functions $\{\nu_i : \Omega_i \to \mathbf{R} : 1 \leqslant i \leqslant N\}$ such that the choice $u_i(t) = c$ is optimal when the states of the processes at time t are $x_i(t)$ $(1 \leqslant i \leqslant N)$ if, and only if,

$$\nu_i[x_i(t)] = \min_{1 \leqslant j \leqslant N} \{\nu_j[x_j(t)]\}.$$

The computation of the functions $\nu_i (1 \leqslant i \leqslant N)$ is facilitated by the following result, due to Nash [4].

Theorem N

If $\nu_i(x_i)$ is defined and finite for all $x_i \in \Omega_i$, then $\nu_i(x_i)$ is the *unique* finite solution of the two equations:

$$\nu_i(x_i) = \frac{E\left\{\sum_{t=0}^{t^*-1} a^t C_i[x_i(t), x_i(t+1)]\right\}}{1 - E(a^{t^*})} \tag{N1}$$

$$t^* = \min_{t>0} \{t \mid \nu_i[x_i(t)] \geqslant \nu_i(x_i)\}. \tag{N2}$$

Intuitively, it may help the reader to think of $\nu_i(x_i)$ as an equivalent stopping cost for the bandit process i in state x_i (if the option of stopping existed).

4. THE SOLUTION

The optimal strategy for the problem outlined in Section 2 is of an especially simple and intuitively appealing form. It is given in the following theorem.

Theorem G

Suppose that at some time, s_i of the n_i phases of project i have been concluded successfully. The optimal strategy is always to allocate research effort to the next phase of the project

(a) which has not failed

(b) which has the smallest associated value of

$$\left\{\sum_{k=s_i}^{n_i-1}\left(\prod_{j=s_i+1}^{k} P_{i,j}\right) C_{i,k+1} - \left(\prod_{j=s_i+1}^{n_i} P_{i,j}\right) R_i\right\}\left\{\prod_{j=s_i+1}^{n_i} P_{i,j}^{-1}\right\}. \quad \text{(G1)}$$

If at any stage adopting this strategy leads to a positive expected cost then research effort should not be allocated to any project.

Proof

Denote by $P(a)$ the problem described in Section 2 in which project i is replaced by the following bandit process ($1 \leqslant i \leqslant m$):

(1) Assume that a phase of the project is completed in unit time. The state of the bandit process is given by:

$$x_i(t) = \begin{cases} -1 \text{ if a phase of project } i \text{ has been unsuccessful before time } t, \\ \text{the number of phases of project } i \text{ which have been successfully} \\ \text{completed, otherwise.} \end{cases}$$

(2) The law of motion of the bandit process is given by:

$$P_i[x_i(t+1) = x_i(t)+1 \mid x_i(t),c] = P_{i,x_i(t)+1}$$

$$P_i[x_i(t+1) = -1 \mid x_i(t),c] = 1 - P_{i,x_i(t)+1} \quad (0 \leqslant x_i(t) < n_i)$$

$$P_i[x_i(t+1) = -1 \mid -1,c] = 1$$

$$P_i[x_i(t+1) = n_i \mid n_i,c] = 1.$$

(3) The process cost function of the bandit process is given by:

$$C_i[-1,-1] = M \text{ for some } M \gg \max_{i,j}(C_{i,j}). \quad \text{(G2)}$$

$$C_i[n_i,n_i] = -R_i(1-a)$$

$$C_i[x_i(t), x_i(t)+1] = C_i[x_i(t), -1] = C_{i,x_i(t)+1} \quad (0 \leqslant x_i(t) < n_i). \quad \text{(G3)}$$

The discount factor is a.

Condition (G2) ensures that in the problem $P(a)$ once a phase of project i has completed unsuccessfully, research effort will be diverted to other projects. If at time t phase n_i of project i completes successfully then clearly from (G3) a total reward of $R_i a^{t+1}$ will accrue from time $t+1$ onwards. Hence $P(1)$ models the problem described in Section 2.

Evidently for $0 \leqslant a < 1$, the optimal strategy for $P(a)$ is given by Theorems GJ and N.

Define the random variables t_i^* for project i ($1 \leqslant i \leqslant m$) as follows:

$t_i^* = j$, if (i,j) is the first phase of project i to be unsuccessful,
∞ if project i is successful.

It can be demonstrated that, for each project i there exists a δ such that whenever

Vol. 4
1976

$a\epsilon[1-\delta,1)$ then t_i^* satisfies both (N1) and (N2) for the bandit process corresponding to project i in problem $P(a)$. It is also true that if project i is such that s_i phases have already concluded successfully then the value $\nu_i(x_i,a)$ which equation (N1) gives when $t^*=t_i^*$ satisfies

$$\lim_{a\to 1} \nu_i(x_i, a) = \left\{ \sum_{k=s_i}^{n_i-1} \left(\prod_{j=s_i+1}^{k} P_{i,j} \right) C_{i,k+1} - \left(\prod_{j=s_i+1}^{n_i} P_{i,j} \right) R_i \right\} \left\{ \prod_{j=s_i+1}^{n_i} P_{i,j}^{-1} \right\}. \quad \text{(G4)}$$

It follows from Theorems GJ and N that the functions $\nu_i(\cdot,a)$ give the optimal strategy for $P(a)$ whenever $a\epsilon[1-\epsilon,1)$ for some ϵ. That the expressions on the right hand side of (G4) give the optimal strategy for $P(1)$ is obvious and can be established by limiting arguments used by Nash [4]. From this follows Theorem G.

Hence, with each of the m projects is associated a profitability index which is merely the expected cost of the project divided by its probability of success. Research effort is applied to the projects beginning with the project with the smallest value of the index. If a project is begun then it is not abandoned until one of its phases is unsuccessful.

5. CONCLUSION

Robert W Cranston, reported on the use of a profitability index in the selection of research projects at B.P's Research Centre, and the advantages which accrued. It is hoped that research managers will use the index given in (G1) as an evaluation of the kind of alternative projects considered here.

It is also my hope that this article will serve in some measure to illuminate the importance of Theorems GJ and N in providing a mathematical basis for project selection.

REFERENCES

1. Cranston RW (1975) First exeriences with a ranking method for portfolio selection. *R&D Mgmt* **5**, 88–92.
2. Gittins JC (1972) Speculative chemical research—How many eggs in a basket? Technical Report CUED/A-Mgt. Stud./TR.4, Cambridge University.
3. Gittins JC and Jones DM (1972) A dynamic allocation index for the sequential design of experiments. *Proc. European Meeting of Statisticians, Budapest.*
4. Nash P (1973) Optimal Allocation of Resources between Research Projects. PhD Thesis, Cambridge University.

ADDRESS FOR CORRESPONDENCE: *KD Glazebrook, Keble College, Oxford OX1 3PG, UK.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

A Deterministic Model for the Housing Decision

RICHARD MOJENA

G GEOFFREY BOOTH

JAMES P BOWMAN

(*Received March 1975; in revised form July 1975*)

The decision to rent or buy housing, or to choose among various housing acquisition alternatives, represents an important and complex financial problem, which confronts virtually all individuals. This paper develops a comprehensive model (within the institutional and legislative context of the United States) which provides theoretically appropriate criteria for evaluating personal housing alternatives. Empirical results clearly indicate the economic consequences of housing alternatives, and both substantiate some and contradict other "wisdoms" concerning the housing decision.

INTRODUCTION

THIS PAPER is concerned with the economic decision of renting vs owning housing for personal use. Within the institutional and legislative context of the United States, a mathematical model is presented which provides the decision maker with a tool for evaluating housing alternatives. The importance of such a model is underscored by noting that over the past two decades in the United States, the average cost of housing amounts to approximately 22% of personal income [8, pp. 322–3]. Furthermore, a commitment to buying housing requires a typical down payment of 20 to 30% of the purchase price, which can vary significantly among regions. For example according to a survey in 1973 [1], median purchase prices were $57,500 for New York City, $34,400 for Chicago, and $53,200 for Los Angeles. Monthly rentals also vary substantially: $398 in New York City, $302 in Chicago, and $240 in Los Angeles.

For the most part, the literature treats the personal housing decision in a descriptive manner, as typified in Refs. [2], [3] and [5]. Existing quantitative approaches lack rigor, fail to explicate relevant variables, ignore the time value of money, or rely on questionable assumptions, as in Refs. [6], [7] and [9]. In

contrast, leasing investment models are well established and documented [4], but inappropriate to the personal housing decision because of differences in the treatment of taxes, depreciation, and capital gains.

The proposed quantitative model explicitly defines theoretically appropriate costs of renting and buying based on a comprehensive set of variables and parameters. By distinguishing between rent–buy aspects, the model may be used to evaluate rent versus buy alternatives, different rentals, or different purchasing options.

Because the pattern of cash flows normally varies among alternatives, differential cash positions result over time, each of which is subject to foregone investment opportunities. Two common methods used to account for this 'time value of money' effect are present value and terminal value. This model employs terminal value because, from an interpretive standpoint, it indicates the total cost of renting or buying up to a given horizon, which includes not only actual costs but the investment income foregone by having to pay for shelter. From another perspective, the terminal value of cost over a given time indicates the amount of funds that an individual would accumulate at the end of the time, including investment income, if payment for housing could be avoided.

In addition, the cost criterion is defined on an after-tax basis, as the incidence of taxes and subsidies varies across housing alternatives. In the United States, for example home owners must pay annual property taxes (typically 1–5% of assessed value) to the local municipality; property taxes and interest payments on the mortgage are deductible for income tax purposes; capital gains tax must be paid to the federal government if a profit is realized on the sale of a house (unless a more expensive purchase is effected within one year); and front-end purchasing subsidies (if any) can be deducted from down payments.

Vol. 4
1976

Since the model excludes qualitative criteria such as quality of life, security of tenancy and other intangible factors, it is appropriately regarded as one of the inputs to the decision-making process. Should it be used as the sole criterion, it must be assumed that qualitative criteria are equal for all alternatives or that alternatives exhibit qualitative balance through tradeoffs.

RENT MODEL

The total cost of renting is defined to consist of the cash flows associated with downpayments, rent payments, and maintenance expenses. The cost of the downpayment, including the impact of the opportunity cost, may be expressed as:

$$D_r \prod_{i=0}^{n-1} [1 + k_i (1 - T_i)] - D_r, \tag{1.1}$$

where

D_r is the initial dollar deposit;
n is the number of periods (months) in the planning horizon;
k_i is the before-tax opportunity cost expressed as a ratio at the *end* of the i^{th} period; and
T_i is the income tax rate in the i^{th} period.

This specification represents the total after-tax income (assuming interest income is taxed at rate T_i) foregone over n periods by paying a rental deposit. This is to say that, if the individual did not have to pay the rental deposit, it could have been invested for n periods at various rates for a net monetary value at the end of the horizon given by (1.1).

Similarly, the cost of rent payments is given by:

$$\sum_{j=1}^{n} \left\{ P_{rj} \prod_{i=j-1}^{n-1} [1 + k_i(1 - T_i)] \right\} \tag{1.2}$$

where P_{rj} is the rent payment in dollars at the *beginning* of the j^{th} period and the remaining notation is as before.

The final category of costs are those associated with maintenance. These include such items as utilities, insurance, and minor repairs not covered by the rental. Additionally, if the geographic locations of housing alternatives vary significantly, then transportation expenses can be included under maintenance. Expressing these costs and their opportunity impact yields:

$$\sum_{j=1}^{n} \left\{ U_{rj} \prod_{i=j}^{n-1} [1 + k_i(1 - T_i)] \right\}, \tag{1.3}$$

where U_{rj} represents maintenance and related costs in dollars at the *end* of the j^{th} period and the other symbols are as previously defined.

Summing these three types of costs and combining certain terms yield the total cost of renting:

$$C_r = D_r \left\{ \prod_{i=0}^{n-1} [1 + k_i(1 - T_i)] - 1 \right\} + \sum_{j=1}^{n} \left\{ P_{rj} \prod_{i=j-1}^{n-1} [1 + k_i(1 - T_i)] \right\} + \sum_{j=1}^{n} \left\{ U_{rj} \prod_{i=j}^{n-1} [1 + k_i(1 - T_i)] \right\}, \tag{1}$$

where C_r is the total cost of renting.

BUY MODEL

The buy model, although similar to the rent model, is somewhat more complex because of additional variables that need to be considered. The total cost of buying is defined to consist of the cash flows associated with the downpayment, principal and interest payments, maintenance expenses, property taxes, and sale of the house.

The time–value cost of the downpayment and related closing costs is given by:

$$(D_b + K) \prod_{i=0}^{n-1} [1 + k_i (1 - T_i)], \tag{2.1}$$

where

D_b is the downpayment in dollars (net of any subsidies); and
K represents closing costs in dollars.

The various periodic costs may be expressed as:

$$\sum_{j=1}^{n} \left\{ [P_{bj} + U_{bj} + (1 - T_j)(I_j + L_j)] \prod_{i=j}^{n-1} [1 + k_i (1 - T_i)] \right\} \tag{2.2}$$

where

P_{bj} is the principal repayment in dollars at the end of the j^{th} period;
U_{bj} represents maintenance and related expenses in dollars at the end of the j^{th} period;
I_j is the interest charge in dollars at the end of the j^{th} period; and
L_j indicates property taxes in dollars at the end of the j^{th} period.

This specification reflects the tax deductibility of interest and property taxes from state and federal income taxes. It should be pointed out that the incidence of the tax deductibility is simplified in this model since it normally does not occur monthly. It should also be noted that the principal repayment and the interest payment are determined from standard annuity formulae based on the amount borrowed, the life of the mortgage, and the mortgage rate.

The last major element in the buy model is that of divesture. The costs associated with this action are:

$$-S_n (1 - b) + [S_n (1 - b) - S_o - K] G + R_n, \tag{2.3}$$

where

S_n and S_o represent the market value of the dwelling in dollars at the end of periods n and o, respectively;
b is the brokerage fee in the form of a ratio;
G is the capital gains tax rate; and
R_n is the remaining principal in dollars.

Vol. 4
1976

The first term represents a cash inflow from selling the house in period n. This amounts to the market value less brokerage fees. The second term is used to calculate the amount of tax owed when the house is sold. This terms should be ignored if it is negative or if another, more expensive, purchase is contemplated within one year. The third term, R_n, represents the amount of the mortgage which has not yet been paid. It is determined easily as a function of the amount borrowed, the mortgage life, the mortgage rate and the length of time the house is owned.

Combining (2.1), (2.2), and (2.3) and simplifying selected terms yield:

$$C_b = (D_b + K) \prod_{i=0}^{n-1} [1 + k_i (1 - T_i)] - S_n (1 - b)(1 - G) - (S_o + K) G + R_n$$

$$+ \sum_{j=1}^{n} \left\{ [P_{bj} + U_{bj} + (1 - T_j)(I_j + L_j)] \prod_{i=j}^{n-1} [1 + k_i (1 - T_i)] \right\}, \qquad (2)$$

where

C_b is the total cost of buying.

CRITERIA

Four criteria of general interest may now be defined: (a) cost of renting for a given planning horizon, (b) cost of buying for a given planning horizon, (c) breakeven period, and (d) cost of buying–renting at breakeven. The first two are determined by the straight-forward calculation of equations (1) and (2), respectively. Note that by defining

$$\Delta C = C_r - C_b,$$

it follows that $\Delta C > 0$ for a given horizon implies a decision in favor of buying, all other things equal, and $\Delta C < 0$ implies a decision to rent.

The breakeven period, n^*, is defined as the first value of n for which $\Delta C \geq 0$, i.e. n^* is the first value of n for which the cost of buying is less than, or equal to, the cost of renting. If the decision maker's expected length of stay is less than the breakeven period, the appropriate financial decision is to rent. The cost of buying–renting at breakeven is obtained by evaluating equation (2) at n^*. As previously expressed, the use of these quantitative criteria must be tempered by qualitative criteria; that is, financial aspects represent only one facet of the housing decision.

DISCUSSION

As might be expected, the rent–buy decision is sensitive to variations in several important variables. The sensitivity of breakeven period is illustrated

TABLE 1. VARIATIONS IN BREAKEVEN PERIOD (MONTHS)

Annual house appreciation rate	Initial Monthly Rental Payments ($)					
	225	250	275	300	325	350
0·00	387	273	185	128	84	59
0·03	214	108	59	40	30	24
0·06	38	28	22	18	15	14

Note: Downpayment set to 25% of $40,000 market price; annual inflation, mortgage, and opportunity cost rates fixed at 0·06, 0·08, and 0·07, respectively; rental deposit assumed equal to initial rent; initial and final income tax rates were 0·20 and 0·25, respectively; U_{ro} = $30/month, U_{bo} = $100/month, K = $300, b = 0·06, t = 300 months, L_0 = $90/month, and G = 0·125; dynamic behaviors of P_{rJ}, U_{rJ}, U_{bJ}, and L_J specified by compounding at rate of inflation.

in Table 1 for variations in the house appreciation rate and the rental payment. The large values of some of the entries clearly dispel the commonly-held belief that 'it is always best to buy housing because equity accumulates.' The decision to rent is favored when the opportunity cost is high relative to the rate of appreciation. Additionally, the effect of lower rental payments is dramatic at low rates of housing appreciation.

To illustrate costs, assume that the alternatives are characterized by a 5 year planning horizon, a 6% annual rate of house appreciation, a monthly rental of $325 per month, and values for other variables and parameters as noted in Table 1. Employing these data, the model indicates that the cost of renting for the 5 year period is $29,512 and the cost of buying for the same period is $18,249. The breakeven period is 15 months and the cost of buying for this period is $5,766. In this case the decision maker should rent if the expected length of stay is less than 15 months. The cost of renting has a unique interpretation: if it were possible to avoid the cost of shelter altogether and cash flows associated with renting were invested at the opportunity cost, the individual would be better off by $29,512 at the end of five years.

Policy implications based on regression analyses of 750 cases include the following: the importance of accurately estimating rates of housing appreciation; high inflation rates favor the purchasing of housing, all other things equal; renting is a favorable financial alternative under a variety of circumstances, particularly for transients and when the opportunity cost is high relative to property appreciation; and the effects of the size of downpayments on the decision to buy or rent and on the costs of buying are minimal, all other things equal.

In concluding, we suggest that this model provides a valid procedure for assessing the financial impact of housing alternatives. As such, it should prove

Vol. 4
1976

useful as one of the inputs to the common and economically important housing decision. In particular, this type of modelling can be used by organizations which provide services related to housing decisions (e.g. realtors and banks) and by personnel departments of organizations which relocate employees (e.g. armed services and large corporations).

REFERENCES

1. BALLABON MB (1974) Inter-city differences in rent vs purchase decisions in housing. *Geography Conference*, West Point, New York.
2. COHEN JB and HANSEN AW (1972) *Personal Finance: Principles and Case Problems*. Richard D Irwin, Homewood, IL.
3. DONALDSON EP and PTAHL JK (1966) *Personal Finance*. Ronald Press, New York.
4. FINDLAY MC III (1974) Financial lease evaluation: survey and synthesis. *Fin. Rev.* 1–15.
5. HASTINGS P and MIETUS N (1972) *Personal Finance*. McGraw-Hill, New York.
6. MUMEY GA (1972) *Personal Economic Planning*. Holt, Rhinehart & Winston, New York.
7. SHELTON JP (1968) The cost of renting vs owning a home. *Land Econ.* **44**, 59–72.
8. U.S. Bureau of the Census (1973) *Stat. Abs. U.S.* Washington, D.C.
9. WEST DA and WOOD GS (1972) *Personal Financial Management*. Houghton Miflin, Boston.

ADDRESS FOR CORRESPONDENCE: *Professor R Mojena, College of Business Administration, Department of Mgmt Science, University of Rhode Island, Kingston RI 02881, USA.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Sampling Information in New Product Marketing

SA CONRAD

(*Received March 1975; in revised form July 1975*)

The traditional evaluation of a new product proposal is a determination of the breakeven volume, with no consideration of the uncertain nature of the data. In this paper sampling information is used to reduce uncertainty. The optimal number of customers to be sampled is determined for a proposal formulated in terms of an expected market share.

INTRODUCTION

THE TRADITIONAL approach to the evaluation of a new product proposal is a determination of the breakeven volume (q_b). Profit (P) is total revenue minus total cost. Total revenue is price (p) times volume (q), and total cost is fixed cost (f) plus the product of unit variable cost (v) and volume. The profit is therefore

$$P = q(p - v) - f,$$

and the breakeven volume is

$$q_b = \frac{f}{p - v}.$$

We shall extend the traditional approach in a simple but useful way, incorporating the uncertain nature of the data, and evaluate a new product proposal formulated in terms of a fixed market size (m) and a market share (θ). We shall assume that each of the m customers either purchases one of the product or does not purchase the product. However, whereas Harris [1] computes the upper bound to the sample size (the m customers may be sampled in an attempt to reduce the uncertain nature of the data), this paper computes the optimal sample size or the optimal number of customers to be sampled.

THE FORMULATION OF THE PROBLEM

Since

$$P = m\theta(p - v) - f$$

then

$$\theta_b = \frac{f}{m(p-v)} \quad \text{and} \quad b = m(p-v),$$

where θ_b is the breakeven market share, and b is the increase in profit per unit increase in θ.

For the two actions

a_1: do not market the product

a_2: market the product

the (opportunity) losses (L) are

$$L(a_1,\theta) = \begin{cases} 0 & \theta \leq \theta_b \\ b(\theta-\theta_b) & \theta \geq \theta_b \end{cases}$$

$$L(a_2,\theta) = \begin{cases} b(\theta_b-\theta) & \theta \leq \theta_b \\ 0 & \theta \geq \theta_b. \end{cases}$$

For example, when $m = 10{,}000$, $p = 15$, $v = 9$, $f = 15{,}000$, then $b = 60{,}000$, $\theta_b = 0{\cdot}25$. The traditional approach is that the company should not market the product if θ is expected to be less than 0·25, and should if θ is expected to be greater than 0·25. We shall introduce uncertainty through a probability distribution for the market share.

Vol. 4
1976

THE MARKET SHARE DISTRIBUTION

The subjective determination of the expected market share is fraught with difficulties and the company should specify a distribution for the market share, θ. The beta distribution

$$f(\theta) = \frac{\Gamma(n)}{\Gamma(r)\,\Gamma(n-r)}\,\theta^{r-1}(1-\theta)^{n-r-1}, \quad (n>r>0),$$

with parameters, r,n, is usually specified. The techniques for the determination of the parameters of the beta distribution are discussed in [2]. If we assume for illustrative purposes that $r = 1$, $n = 5$, then the expected market share is

$$E_0 = \frac{r}{n} (= 0{\cdot}20),$$

and therefore the company should not market the product. (The subscript 0 denotes expectation *without* further information.)

The further information is sampling information. If the company now samples N customers, and observes R customers purchasing (or intending to purchase) the product, then the distribution for θ is a beta distribution with

parameters $r + R$, $n + N$. Therefore

$$E_N = \frac{r+R}{n+N}$$

and the company's action changes from 'do not market' to 'market' if

$$\frac{r+R}{n+N} > \theta_b$$

or

$$R > \theta_b(n+N) - r. \tag{1}$$

(The subscript N denotes expectation after sampling N customers.) This sampling information reduces the expected loss by $b(E_N - \theta_b)$ for the values of R in (1); the optimal number of customers to be sampled is the number corresponding to the maximum net expected reduction in loss.

THE OPTIMAL SAMPLE SIZE

The probability of this reduction is

$$\frac{\Gamma(n)\Gamma(N+1)\,\Gamma(r+R)\,\Gamma(n+N-r-R)}{\Gamma(r)\,\Gamma(n-r)\,\Gamma(N-R+1)\,\Gamma(R+1)\,\Gamma(n+N)}$$

and the expected reduction in loss is therefore

$$L(N) = \sum_{R=R_B(N)}^{N} b\left(\frac{r+R}{n+N} - \theta_b\right) \frac{\Gamma(n)\,\Gamma(N+1)\,\Gamma(r+R)\,\Gamma(n+N-r-R)}{\Gamma(r)\,\Gamma(n-r)\,\Gamma(N-R+1)\,\Gamma(R+1)\,\Gamma(n+N)}$$

where $R_B(N)$ is the smallest value of R that changes the action from 'do not market' to 'market' and $N = 1,2, \ldots$. The net expected reduction in loss is

$$S(N) = L(N) - K(N)$$

where $K(N)$ is the sampling cost.

If we assume $K(N)$ to be linear and in particular equal to $500 + 40N$, then the company should sample 13 customers (see Fig. 1). If less than 4 purchase (or intend purchasing) the product, the company should not market the product. If 4 or more, the company should market the product. The expected profit increases from 0 (without sampling information) to 1283.

CONCLUSION

Sampling information is used in an attempt to reduce the uncertain nature of the data in the evaluation of a new product proposal. The paper computes the

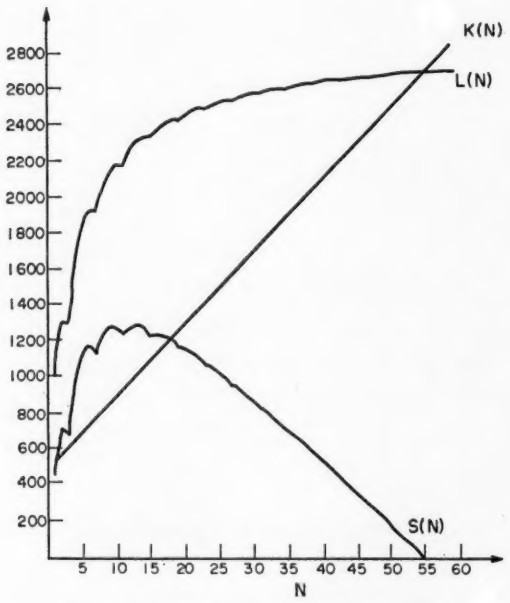


FIG. 1. *The net expected reduction in loss.*

optimal number of customers to be sampled for a proposal formulated in terms of a fixed market size and a market share.

It is clear that marketing mix elements will have an effect on the market share. The author is investigating the price element by the specification of different beta distributions for different prices. The different prices are then compared on a profit-uncertainty grid, rather than the deterministic optimisation of the market share as a function of price. The author is also analysing the problem when, in general, there are less than N responses from the N customers.

REFERENCES

1. HARRIS L (1966) New product marketing: a case study in decision-making under uncertainty. *J.R. Statist. Soc. C.* **15**, 39–42.
2. WINKLER RL (1967) The assessment of prior distributions in bayesian analysis. *J. Am. Statist. Ass.* **62**, 776–800.

ADDRESS FOR CORRESPONDENCE: *SA Conrad, Department of Management Science, UMIST, POB88, Manchester M60 1OD, UK.*

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Memoranda

Bayesian Point Estimation on the Bernoulli Parameter

THE FULL theory of Bayesian point estimation Lavalle [7] requires the individual to select as his estimate that particular value which maximises his expected utility. There are many situations, however, where although the Bayesian facility to define a probability measure on the set of hypotheses is theoretically attractive, the introduction of a personal value function on the realisation of these hypotheses may not be appropriate. The scientific model discrimination scheme of Zellner [12] in econometrics is such a case. In other instances, the organisational context may be such that the full decision analysis and assessment of utility is separated from the individual technical "expert" who formulates the estimate. Many forecasters have to provide "good" point estimates without even knowing the contingent decisions which may be taken.

With regard to this problem, there appears to be some controversy in the Bayesian literature upon whether the posterior mean or mode is the "better" predictor. Some forecasters argue that the posterior mode is a more sensible predictor. Using a maximum likelihood argument, Lee, Judge and Zellner [8] for example, advocate the use of the posterior mode as a point estimate for the transition probabilities in a Markov process. However, from a stricter Bayesian point of view, the posterior mean is the Bayesian expectation and as such would be adopted by followers of de Finetti [4]. Good [5] in the context of multinomial estimation strongly advocates the posterior mean on account of its being the "final Type II expectation of p_i and the final Type II probability that the next item will be of the i^{th} category".

In the context of orthodox point estimation, the posterior mode is analogous to a maximum likelihood estimator and the posterior mean to an unbiased estimate. Pratt [1] and Lavalle [7] have examined the conditions of sufficiency for such equivalence. They appear to reduce to the accuracy of the prior density function with respect to the sample of data. For equivalence, Lavalle [7] requires that the sample outcome should be 'unsurprising'. This is not very helpful, however, in trying to decide *a priori* whether the posterior mean or mode is the more appropriate predictor. Consider, for example, the Bayesian analysis of the beta process, which is finding increasing applications in management science [e.g. 2, 3, 10].

If k is distributed $B(k|a_1, a_2)$ the density function

$$p(k) = [B(a_1, a_2)]^{-1}\, k^{a_1-1} (1-k)^{a_2-1},$$

where

$$B(a_1, a_2) = \Gamma a_1\, \Gamma a_2/\Gamma(a_1 + a_2)$$

and for k

$$\text{mean} = a_1/(a_1 + a_2)$$

$$\text{mode} = (a_1 - 1)/(a_1 + a_2 - 2)$$

$$\text{variance} = a_1 a_2/(a_1 + a_2)^2(a_1 + a_2 + 1)$$

with $k \in (0,1); 0 < a_1; 0 < a_2$.

In the forecasting context, a new item of data, δ, will represent whether F_1 (forecasting model 1) has outperformed F_2, when $\delta = 1$; or not, when $\delta = 0$. Hence δ can be considered to be a Bernoulli variable. After j forecast realisations, and providing the forecaster believes the *pdf* to be stationary, his posterior distribution for k should be

$$B(k \mid a_1 + s_j, a_2 + j - s_j),$$

where

$$s_j = \sum_{i=1}^{j} \delta_i$$

The posterior mean is given simply from the

formula

$$\overline{k} = (a_1 + s_j) / (a_1 + a_2 + j)$$

This can be re-written as $\overline{k} = \lambda P + (1 - \lambda) D$, where $P = a_1/(a_1 + a_2)$, the prior mean, $D = s_j/j$, the sample mean and maximum likelihood estimator of k and $\lambda = (a_1 + a_2)/(a_1 + a_2 + j)$, the linear weighting factor. Clearly as j increases, the proportionate weight placed on prior belief decreases, but never vanishes.

With the usual diffuse prior $a_1 = a_2 = 1$ to represent complete uncertainty, it is clear that the posterior mode becomes identical to the unbiased maximum likelihood estimator of k. In other words, given such a diffuse prior, the posterior mode gives a point estimate which can be considered to be based only upon the sample data. Whilst this may seem reasonable for posterior belief based upon large stationary processes (given that we have no prior knowledge on the process, why should we dilute experimental evidence with two extra dichotomous outcomes?), it does not seem so appropriate to small samples. Moreover, it is precisely for small samples, i.e. relatively diffuse *pdfs*, that the problem is important. Diffuse *pdfs* are the more common situation as very few processes are stationary over enough realisations for the *pdf* to become highly informative.

With the $B(k|1,1)$ diffuse prior, the expectation for the first observation is $P(F_1) = P(F_2) = 0{\cdot}5$ which is reasonable as equal weighting, but the mode is undefined. After one realisation, say F_1 outperforms F_2, then the posterior expectation that F_1 will outperform F_2 next time is $P(F_1) = 2/3$ which is more intuitively acceptable than the posterior mode $P(F_1) = 1$, which implies certainty. Indeed, it is not just this pragmatic judgement which militates against the mode, it is this philosophically untenable propensity to give probabilities zero or unity to inductive hypotheses. Even orthodox statisticians, after observing $s_j = j$, would mostly be unwilling to bet everything on the certainty of the next observation.

It could be argued, however, that, in the decision-neutral context of scientific investigation, the posterior mode, being a purely empirical measure of the relative support for each hypothesis, could be defended as the more appropriate estimate. Nevertheless, it is suggested that the inferential basis for an empirical Bayesian analysis is fundamentally different to that of the pragmatic decision-theoretician.

Basic to the decision-theoretic analysis of the Bernoulli process is the $B(k|1,1)$ uniform prior to represent the initial complete uncertainty. This follows Bayes and Laplace and is generally advocated on pragmatic grounds for decision-making under uncertainty (cf. Aitchison [1]). Its justification lies in the Principle of Insufficient Reason whereby each possible hypothesis is assigned equal probability. However, because of the way in which the beta distribution is formulated, it is open to the criticism of introducing two artificial bits of information into the analysis.

A purely empirical Bayesian analysis would therefore seek to eliminate this by allowing a_1 and a_2 in $B(k|a_1,a_2)$ to approach zero. This gives the U-shaped improper prior advocated by Lindley [9].

$$P(k) = 1/k(1-k).$$

Jeffreys [6] also defends this U-shaped prior, which he refers to as Haldane, on the grounds that it can provide an innocuous initial limiting distribution from which to build a Bayesian analysis of the Bernoulli process.

A consequence of the use of this Haldane prior, however, is that the posterior mean is now identical to the posterior mode if the analysis had been built upon a uniform prior. Thus, any attempt to extend this measure for inductive or future-contingent decision purposes exposes it to exactly the same criticism as the posterior mode under the uniform prior. It should be emphasised that any justification for the use of the Haldane prior can therefore be based only upon its facility to provide a purely empirical *descriptive* probability measure of past performance and experimental evidence. Its unsuitability for predictive and decision analyses has already been indicated.

Thus the case for using the posterior mean as a point estimate is generally a strong one and the source of controversy is apparently the more fundamental question of recognising the precise type of probability measure which is required by the analysis.

REFERENCES

1. Aitchison J (1970) *Choice Against Chance*. Addison-Wesley, New York.

Vol. 4
1976

2. BIERMAN H and HAUSMAN WH (1970) The Credit Granting Decision. *Mgmt Sci.* **16B,** 519–532.
3. BUNN DW (1975) A Bayesian Approach to the Linear Combination of Forecasts. *Opl. Res. Q.* **26,** 325–329.
4. DE FINETTI B (1974) *Theory of Probability.* Wiley, New York.
5. GOOD IJ (1967) A Bayesian Significance Test for Multinomial Distributions. *J. R. Statist. Soc.* **29B,** 399–431.
6. JEFFREYS H (1961) *Theory of Probability.* Oxford Univ. Press.
7. LAVALLE IJ (1970) *An Introduction to Probability, Decision and Inference.* Holt, Rinehart & Winston, New York.
8. LEE TC, JUDGE GG and ZELLNER A (1968) Maximum Likelihood and Bayesian Estimation of Transition Probabilities *Am. Statist. Assoc. J.* **63,** 1162–1179.
9. LINDLEY DV (1965) *Probability and Statistics.* Cambridge Univ. Press, London.
10. MURRAY GR and SILVER EA (1966) A Bayesian Analysis of the Style Goods Inventory Problem. *Mgmt Sci.* **12,** 785–797.
11. PRATT JW (1965) Bayesian Interpretation of Standard Inference Statements. *J. R. Statist. Soc.* **27B,** 169–203.
12. ZELLNER A (1971) *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.

DW Bunn
(*October 1975*)

The London Graduate School of Business Studies
Sussex Place
Regent's Park
London NW1 4SA

ol. 4
976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Some Numerical Equilibrium Results for the D/G/1: (∞/FIFO) Queueing Model

AS THE solution of even conceptually simple queueing models is in general difficult, requiring mathematical skills of a high standard, long learning and understanding processes, access to computer facilities and in many cases a considerable amount of computer time, the availability of charts or tables providing numerical results for measures of practical interest is highly desirable.

The model reflecting the actual waiting line situation under study need not necessarily match one of the models for which numerical results are available. Such results may either be viewed as rough estimates of the true values if extreme accuracy is not required, or as bounds on the relevant values if a "conservative" or "worst case" type of analysis is considered as adequate.

This note gives some numerical equilibrium results for the average queueing time of an arriving customer (W) and the probability of non zero queueing time in models of the general type D/G/1:(∞/FIFO) as defined by Lee's [4] notation.

TABLE 1. μW OBTAINED ANALYTICALLY FOR SOME $D/E_k/1$:(∞/FIFO) MODELS

ρ \ k	1	2	3	4	6	10
0·50	0·255001	0·078415	0·034185	0·017306	0·005532	0·000837
0·55	0·352815	0·119473	0·056946	0·031430	0·011909	0·002535
0·60	0·379822	0·175577	0·089769	0·052967	0·022834	0·006264
0·65	0·647715	0·252581	0·136636	0·084992	0·040501	0·013460
0·70	0·876160	0·360305	0·204136	0·132511	0·068360	0·026338
0·75	1·200729	0·516513	0·304137	0·204464	0·112459	0·048663
0·80	1·692731	0·756831	0·460398	0·318699	0·184758	0·087724
0·85	2·518735	1·164514	0·728427	0·516858	0·313063	0·160295
0·90	4·178659	1·989492	1·274838	0·923913	0·580716	0·316452
0·95	9·172434	4·481704	2·932918	2·164831	1·404381	0·806241

TABLE 2. P OBTAINED ANALYTICALLY FOR SOME $D/E_k/1$:(∞/FIFO) MODELS

ρ \ k	1	2	3	4	6	10
0·50	0·203188	0·116449	0·072619	0·047278	0·021496	0·005073
0·55	0·260801	0·165813	0·113513	0·080763	0·043615	0·014461
0·60	0·324243	0·224455	0·165585	0·126336	0·077905	0·033350
0·65	0·393099	0·292199	0·229191	0·185004	0·126702	0·066004
0·70	0·466996	0·368816	0·304506	0·257486	0·191966	0·116694
0·75	0·545605	0·454047	0·391580	0·344277	0·275308	0·189462
0·80	0·628630	0·547623	0·490371	0·445700	0·378026	0·288003
0·85	0·715807	0·649271	0·600774	0·561941	0·501150	0·415626
0·90	0·806900	0·758723	0·722637	0·693082	0·645481	0·575233
0·95	0·901695	0·875717	0·855780	0·839123	0·811620	0·769323

Tables 1 and 2 correspond to the case where service times have the Erlang distribution given by:

$$dB(t) = \frac{(k\mu)^k t^{k-1} \exp(-k\mu t)}{(k-1)!} dt$$

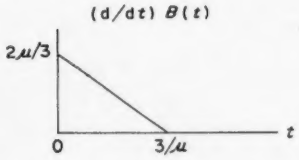
with mean $1/\mu$, variance $1/(k\mu^2)$ and coefficient of variation $v_s = \sqrt{1/k}$. For this case, the analytical method of solution discussed by Cox and Smith [1] was adopted and applied for $k = 1, 2, 3, 4, 6$ and 10.

Table 3 is related to the three distributions of service times:

TABLE 3. μW AND P OBTAINED BY SIMULATION FOR SOME $D/G/1:(\infty/\text{FIFO})$ MODELS

Model description			Results of simulation runs					
			ηW			P		
$B(t)$	v_s^2	ρ	lower 95% conf. limit	sim/ed	upper 95% conf. limit	lower 95% conf. limit	sim/ed	upper 95% conf. limit
S.T.	1/2	0·75	0·4675	0·4769	0·4864	0·4871	0·4895	0·4919
S.T.	1/2	0·80	0·6992	0·7164	0·7336	0·5782	0·5805	0·5828
S.T.	1/2	0·85	1·0878	1·1213	1·1548	0·6777	0·6804	0·6831
S.T.	1/2	0·90	1·8953	1·9510	2·0067	0·7805	0·7823	0·7841
R	1/3	0·80	0·3942	0·4026	0·4110	0·5431	0·5453	0·5474
R	1/3	0·85	0·6522	0·6691	0·6860	0·6483	0·6501	0·6518
R	1/3	0·90	1·1873	1·2162	1·2451	0·7603	0·7616	0·7630
R	1/3	0·95	2·7447	2·8600	2·9752	0·8775	0·8789	0·8804
D.T.	1/6	0·85	0·2741	0·2791	0·2841	0·5244	0·5265	0·5287
D.T.	1/6	0·90	0·5323	0·5461	0·5599	0·6700	0·6719	0·6738
D.T.	1/6	0·95	1·3194	1·3586	1·3977	0·8288	0·8304	0·8320

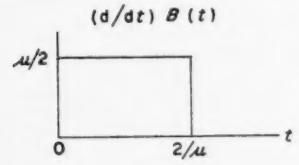
1. *Simple triangular* (S.T.)



$$dB(t) = \begin{cases} (2\mu/9)/(3-\mu t)\,dt & \text{if } 0<t<3/\mu \\ 0 & \text{otherwise} \end{cases}$$

with coefficient of variation v_s equal to $\sqrt{1/2}$;

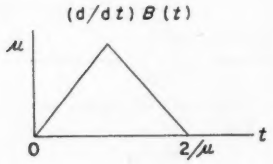
2. *Rectangular* (R)



$$dB(t) = \begin{cases} (\mu/2)dt & \text{if } 0<t<2/\mu \\ 0 & \text{otherwise} \end{cases}$$

with coefficient of variation v_s equal to $\sqrt{1/3}$; and

3. *Double triangular* (D.T.)



$$dB(t) = \begin{cases} \mu^2 t\, dt & \text{if } 0<t<1/\mu \\ \mu(2-\mu t)\, dt & \text{if } 1/\mu<t<2/\mu \\ 0 & \text{otherwise} \end{cases}$$

with coefficient of variation v_s equal to $\sqrt{1/6}$.

For the three cases considered in Table 3, estimates of the average queueing time and of the probability of non zero queueing time were obtained by simulation using the SIMON language [2] and the Imperial College CDC 6400 computer. In each particular run 200,000 arrivals were simulated (the requirements in computer time being equivalent to approx. 8 min) and variance reduction techniques were applied for the calculation of 95% confidence limits on the simulated values.

According to Kingman [3], under saturated conditions ($\rho \to 1$) the average queueing time of an arriving customer to a GI/G/1:(∞/FIFO) system is a function of ρ and of the coefficients of variation of inter-arrival and service times only. A comparison, however, between Tables 1 and 3 seems to indicate that, at least for the queueing system considered in this note, the result established by Kingman may possibly be approximately true for values of ρ encountered in many practical situations. Hence the question: When dealing with D/G/1:(∞/FIFO) models and interested in estimating W, could one use the values presented in Table 1 regardless of the particular distribution of service times if the actual μW to be estimated is not too small? Not smaller than, say, 0·5?

REFERENCES

1. Cox DR and Smith WL (1961) *Queues*. Methuen, London.
2. Hills PR and Mathewson SC (1972) A programming manual for SIMON (in FORTRAN), Department of Management Science, Imperial College, London.
3. Kingman JFC (1962) On queues in heavy traffic. *J. R. Statist. Soc. B* **24** 383.
4. Lee AM (1966) *Applied Queueing Theory*. St. Martin's Press, Montreal, Canada.

George P Cosmetatos
(*September 1975*)

Department of Management Science
Imperial College of Science and Technology
Exhibition Road
London SW7 2BX

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976, Pergamon Press. Printed in Great Britain

Macro-Efficiency in Interactive Work Units

MANAGEMENT provides inputs to a group of employees and expects finished outputs from the group. Traditionally, time and motion study analyzes time inputs by reducing work tasks to precise micro-elements. By its nature, the approach ignores the macro input–output relationships of the work unit, especially worker interdependence. In work units where there is a discernible degree of interdependence among employees, 'real' productive output time is not necessarily the time spent on the job. Aside from necessary time lost for vacations, holidays, and sick leave, there is 'interaction time' between workers. Such times do not *directly* contribute to physical or 'external' output, so that micro concepts of efficiency cannot properly account for such interdependencies. This note shows how the Simpson [2] organizational input–output model can supplement micro time and motion study by providing decision information about the macro efficiency of interactive work units.

In Simpson's model, productive or 'real' working time is that spent in efforts directly related to the tasks of physically producing the output of the work unit. 'Real' time is only one component of 'actual 'time; others are personal delays and relief, supervision of workers, management time charges (time demanded of the worker by management and delays which are part of working conditions), and productive interaction among employees (time spent by each worker in helping one another in the performance of tasks).

Columns in Simpson's input–output table represent inputs of time by each worker, and rows represent outputs of time from each worker. Macro efficiency of the organization of activities in the work unit is given by the Niebel ratio (1, page 508], which is the percentage of productive effort that results from each clock hour of operation. In terms of the input–output model, the Niebel ratio is:

$$\text{NR} = \frac{\text{total external output time}}{\text{total actual output time}}.$$

Other measures of macro efficiency available from the model include: (1) the Leontief inverse—whose column sums give the total hours of actual output time expended by each worker to obtain one hour of external output time, or each worker's efficiency within the work unit, and (2) the disaggregated total output matrix—whose column sums can be divided by the disaggregated total external vector column sums to form the opportunity cost in terms of time for each work operation, giving the efficiency of each work operation within the work unit.

The methodology can be illustrated by summarizing the results of a case study in Table 1, collected at nominal cost primarily by a structured questionnaire. In Table 1 the supervisor column represents the demand made upon the group through the supervisor for management time charges as follows: 14 = the time taken by the supervisor for personal delays (including his response to higher level management) and relief; 1·8, 4·7, 2·1 = management time charges of workers 1, 2, and 3, respectively.

Each worker's column shows an input of supervision, the worker's personal delays and relief, and any input of help from other workers. For example, worker 1's column shows: 2·0 = input of the supervisor's time to this worker; 3·3 = personal delays and relief time of this worker; 2·9 and 1·0 = inputs of time from workers 2 and 3, respectively, to help worker 1.

TABLE 1. COMPANY ADMINISTRATIVE SERVICES INPUT–OUTPUT MATRIX
(hr per standard 40-hr week)

Outputs \ Inputs	Supervisor	Worker 1	Worker 2	Worker 3	External Output	Total Output
Supervisor	14·0	2·0	2·0	2·0	20·0	40
Worker 1	1·8	3·3	10·0	10·0	14·9	40
Worker 2	4·7	2·9	1·0	3·5	27·9	40
Worker 3	2·1	1·0	2·5	2·0	32·4	40
External Input	17·4	30·8	24·5	22·5	95·2	
Total Input	40	40	40	40	40	160

Entries in the external input row represent the time allocated to each worker for his effort, excluding the internal inputs received from supervision and co-workers. For example, the 17·4 of external input for the supervisor is the difference between his total input of 40 and the sum of his internal inputs of 22·6 (14·0 + 1·8 + 4·7 + 2·1). Entries in the external output column represent the 'real' or productive output of the work unit; they are the difference between the total output in each row (40 for the supervisor as an example) and the sum of the internal outputs in that row (20 = 14·0 + 2·0 + 2·0 + 2·0 for the supervisor).

The Niebel ratio is NR = 95·2/160 = 0·595, meaning that 0·595 hr of productive effort result from each weekly clock hour of operation, or it requires 1·68 hr of paid weekly clock time to obtain one hour of productive working time. When the time lost for vacations, holidays, and sick leave is removed from the annual work year, the annual Niebel ratio is NR = 4454/8320 = 0·535, indicating that these fringe benefits result in a 10·1% opportunity cost in 'real' or productive time output efficiency.

Calculation of the Leontief inverse (not reported here for the sake of brevity) for this work unit indicated that the supervisor requires 2·054, worker one 1·3736, worker two 1·591, and worker three 1·668 clock hours to obtain one hour of 'external or real' time output. Therefore, the supervisor is the least efficient and worker one the most efficient worker in terms of time. This result is not surprising, since, in this work unit, the supervisor spends a large portion of his time in physical production while worker one acts as a *de facto* supervisor or 'lead worker' supervising the work of others.

The case presented here represents the first application of this methodology, so that no other empirical macro NR values are available for comparison. An *a priori* upper bound of NR = 1·000 can be established from the model's construction; it is the lower bound which presents the difficulty. In theory, it is possible that the lower bound may be an NR = 0, but in practice it is unlikely that we would observe many NR < 0·50. In another case application we found the weekly Niebel ratio NR = 0·787, and the annual NR = 0·708.

Further research seems warranted, especially in establishing empirical NR values for various types of work units

REFERENCES

1. Niebel Benjamin W (1967) *Motion and Time Study*. Irwin, Homewood, IL.
2. Simpson TD (1973) An Application of Input–Output Analysis to Staffing Problems. Unpublished M.B.A. thesis, California State Polytechnic University, Pomona, California.

Cuissepi A Forgionne
Truman D Simpson
(*August 1975*)

Business Management Department
School of Business Administration
California Polytechnic University
Pomona
California 91768
U.S.A.

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

The Multiproportional RAS Method

THE MULTIPROPORTIONAL RAS method has been designed to eliminate inconsistencies between the long-term plan and the short-term budget of an organization. Why do such inconsistencies arise? Because the way in which the long-term plan is compiled is the reverse in two respects of the way in which the budget is compiled.

The long-term plan is compiled top-down and from the future backwards. First, global targets for the organization are set far ahead. This results in an aggregate strategic long-term plan which is subsequently broken down and extended backwards until "next year", i.e. the first year of the plan.

The short-term budget is compiled bottom-up and from the past forwards. For instance, a sales manager will first consider how individual sales items (possibly down to individual clients) have been in the past, especially last year, and extrapolate into next year taking account of all available additional information. The detailed figures for next year are subsequently aggregated until the complete draft budget with gross sales and profits or losses emerges.

As a rule, the first year of the long-term plan and the draft budget are not consistent. However, both have their merits. We will assume here that the first year of the plan gives the best possible estimates of the more aggregate figures whereas the draft budget gives the best possible estimates of the detailed pattern of individual budget or sales items (in the following, we shall concentrate on sales). However, the first year of the plan does not give the detail required for the budget and the draft budget tends to be afflicted with systematic errors. The systematic errors may be underestimations: a sales manager tends to have more information about old sales opportunities or old clients that will disappear than about new sales opportunities or new clients that will emerge. Or they may be overestimations: a sales manager in an optimistic mood may inadvertently put down high individual sales figures which only after summation will prove to be unrealistically high.

The inconsistencies may be eliminated by the multi-proportional RAS method which is a generalization of the (bi-proportional) RAS method proposed by Stone and Brown [2]. In the bi-proportional RAS method, a given matrix, A, should be made consistent with given column and row totals. It is assumed that there are systematic errors such that all elements of a row should be proportionally corrected by a common row factor and that all elements of a column should be proportionally corrected by a common column factor. Hence, any one element of the A-matrix is simultaneously subject to the correction factors of the row and the column to which it belongs. The corrected matrix, A*, can be written as the matrix product RAS (hence the name) of the original matrix, A, pre-multiplied by a diagonal matrix, R, of row correction factors and post-multiplied by a diagonal matrix, S, of column correction factors. The correction factors are computed by an iterative algorithm. It has been shown by Bacharach [1] that the algorithm converges under weak assumptions and that the result has some attractive properties, e.g. zero elements of the original matrix are preserved and the corrected matrix is, in a sense, near to the original matrix.

The RAS method can be generalized to solve our present problem. Let budgeted sales be characterized by, say, four dimensions. For instance, a sales item may be specified by (i) product, (ii) location where it is produced, (iii) market, (iv) period—month or quarter—in which it is produced. The draft budget will thus consist of a four-dimensional array, A(H,I,J,K), where the first index (H) indicates the product, the second (I) the location, the third (J) the market, and the fourth (K) the period. Let the first year of the long-term plan only specify marginal totals: by products, by locations, by markets, and by periods. It is assumed that there are systematic errors such that all elements of the array should be proportionally corrected by common factors for products, locations, markets, and periods in order to make the budget consistent with the long-term plan. Thus, four correction factors act simultaneously on any one element of the array. The method of computing the multiproportional correction factors is analogous to the bi-proportional RAS method. It is presented in [3]. Although convergence remains to be proved, the

algorithm has worked, so far, under diverse conditions.

The multi-proportional RAS method is not as mechanical as it may look. It may be used to eliminate remaining inconsistencies after all prior information has been exhausted. Also, individual items may be corrected by hand and exempted from the RAS procedure. Owing to the zero-preserving property, they can be taken out of the array and subtracted from the marginal totals before the procedure and can be inserted again afterwards. It is, of course, a necessary condition for a useful application of the multi-proportional RAS method that the budgeting activity is processed on a computer.

REFERENCES

1. BACHARACH M (1970) *Biproportional Matrices and Input–Output Change*. Cambridge University Press, London.
2. STONE R and BROWN JAC (1962) A Long-Term Growth Model for the British Economy. In: RC Geary (editor) *Europe's Future in Figures*. North-Holland, Amsterdam.
3. TILANUS CB (1974) Where Short-Term Budget Meets Long-Term Plan: Solving Conflicts by the Multiproportional RAS Method. Paper presented at the VIth International Conference on Input–Output Techniques, Vienna, 22–26 April 1974. To be published in: CB Tilanus (editor) *Quantitative Methods in Budgeting*. Rotterdam University Press, Rotterdam.

CB Tilanus
(*September 1975*)
Department of Industrial Engineering
University of Technology
PO Box 513
Eindhoven
Netherlands

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Variance Analysis Revisited

A VARIANCE in accounting differs from the statistical variance concept in that it is simply defined as the difference between two amounts, e.g. the difference between a budgeted cost item and the actual item. Variance analysis is concerned with breaking down variances into meaningful parts.

What is a meaningful part of a variance? The answer can be derived from responsibility accounting and flexible budgeting. Responsibility accounting tries explicitly to allocate budgeted costs and their variances to those managers in the organization who can exert a predominant influence on the costs and can thus be held responsible for them. In flexible budgeting, costs are not budgeted as fixed amounts or fixed rates but rather as functions.

Cost items can quite generally be viewed as functions. For instance, in the case of a cost item of two dimensions, total cost is a function of quantity produced. Even if three categories of costs are distinguished, viz.

(i) fixed costs
(ii) proportionally variable costs
(iii) 'semi-variable' costs

(where the latter category is any arbitrary cost function), the categories (i) and (ii) can be considered as special cases (hyperbola and straight line, respectively) of (iii). In a function of two variables, one (dependent) variable is a function of another (argument) variable. The dependent variable is the effect, the argument variable is the cause. In the case of a cost item of three dimensions, total cost is a function of input volume, which in turn is a function of quantity produced. In the three-dimensional case, we have a recursive cost model which can be sequentially solved: if a numerical value is substituted for quantity produced, input volume can be computed; next, if (computed) input volume is substituted, total cost can be computed.

The central concept of this article is that the chain of cause and effect determines responsibility and provides the clue for a meaningful break-down of variances. Thus, if actual quantity produced deviates from budgeted quantity produced, this entails functional deviations of input volume and total cost due to the flexible budget functions. All deviations, both direct and indirect, should be attributed to the person responsible for the quantity produced, e.g. the sales manager. Next, if actual input volume deviates from budgeted input volume computed as the value of the budget function for actual quantity produced, this deviation plus the entailed indirect effect on total cost should be attributed to the person responsible for the input volume, e.g. the production manager. Finally, if actual total cost deviates from budgeted total cost computed as the value of the budget function for actual quantity produced and actual input volume, this deviation should be attributed to the person responsible for the procurement of input, e.g. the purchasing manager.

Furthermore, there is a time dimension. One may distinguish long-term "normal" conditions on which long-term planning is based and short-term business conditions on which the budget is based. It is assumed that different persons in the organization are responsible for deviations from long-term standards (e.g. R&D managers) and from short-term standards (e.g. Operations managers).

The computational scheme for a two-dimensional, two-period break-down is given below. The geometrical meaning of the break-down may be traced in Fig. 1.

The two-dimensional, two-period case is just one of the more practical examples of the method of variance analysis consistent with flexible budgeting and responsibility accounting. Theoretically, the method is valid for any number of dimensions and periods.

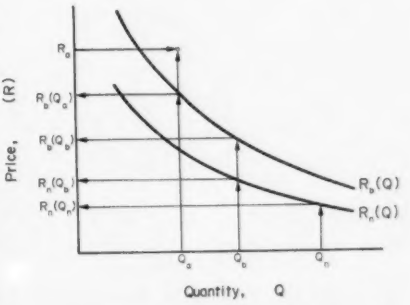


FIG. 1. *Two-dimensional, two-period break-down of total variance.*

$$
\begin{aligned}
V_{n,a} &= C_n - C_a && \text{total variance} \\
&= Q_n + R_n(Q_n) - Q_a \times R_a \\
&= Q_n \times R_n(Q_n) - Q_b \times R_b(Q_b) && \text{long-term variance} \\
&\quad + Q_b \times R_b(Q_b) - Q_a \times R_a && \text{short-term variance} \\
&= [Q_n - Q_b] \times R_n(Q_n) && \text{direct} \\
&\quad + Q_b \times [R_n(Q_n) - R_n(Q_b)] && \text{indirect} \Big\} \text{effect of deviation } Q_n - Q_b \\
&\quad + Q_b \times [R_n(Q_b) - R_b(Q_b)] && \text{effect of deviation } R_n - R_b \\
&\quad + [Q_b - Q_a] \times R_b(Q_b) && \text{direct} \\
&\quad + Q_a \times [R_b(Q_b) - R_b(Q_a)] && \text{indirect} \Big\} \text{effect of deviation } Q_b - Q_a \\
&\quad + Q_a \times [R_b(Q_a) - R_a] && \text{effect of deviation } R_b - R_a
\end{aligned}
$$

where

n = long-term, 'normal'
b = short-term, 'budgeted'
a = actual
V = variance
C = total cost (of any given cost item)
Q = quantity produced (volume)
R = price (per unit of quantity produced).

More extensive treatments of the method are given in [1, 2]. The reader may make sure for himself that any other method of variance analysis that may be encountered in the literature seems illogical.

REFERENCES

1. TILANUS CB and THEEUWES JAM (1975) Variance Analysis Revisited. Paper presented at EURO 1, 1st European Conference of Operational Research, Brussels, January 1975.

2. TILANUS CB and THEEUWES JAM (1975) Variance Analysis, Flexible Budgeting and Responsibility Accounting. (Ed. Tilanus CB) *Quantitative Methods in Budgeting*. Rotterdam University Press, Rotterdam (forthcoming).

CB Tilanus
JAM Theeuwes
(*September 1975*)

Department of Industrial Engineering
University of Technology
PO Box 513
Eindhoven
Netherlands

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

Alternative Estimators of Lines of Securities

THE CHARACTERISTIC line of a security or portfolio relates its rate of return to that of the market portfolio. Several normative portfolio selection models require the slope of this characteristic line for individual securities as an input to select *ex ante* efficient portfolios [5, 6]. One possible approach to the measurement of this parameter is the regression of historical returns of the securities on those of an index chosen to approximate the market portfolio. Since there is evidence which suggests that the distribution of returns may be represented by a member of the Stable family with undefined variance [4], it has been suggested that the regression satisfy the criterion of minimizing the sum of absolute errors rather than the sum of squared errors around the regression line [3]. It has been alleged that the differences in the estimates of regression coefficients so obtained is inessential, at least for well diversified portfolios, and thus it is concluded that the choice of estimators is immaterial for any purpose [7]. No attempt is made, however, to proceed beyond the comparison of regression coefficients for individual securities to examine their implications for the *ex post* results of choosing *ex ante* efficient portfolios using the alternative estimates as inputs. This note presents such a comparison and observes that marked differences are observed.

The comparison is based on the data and method of Cohen and Pogue [1], extending the set of models to include the linear portfolio model of Sharpe [6] using both ordinary least squares (OLS) and minimum sum of absolute error (MSAE) regression estimates as input data. The returns data from a random selection of 75 securities were divided into an *ex ante* period for parameter estimation, and an *ex post* period for evaluation [1, pp. 173–5]. Two additional sets of *ex ante* efficient portfolios were generated using the constraint that each portfolio contain a minimum of 20 securities,[1] and having the same *ex ante* returns equally spaced between the maximum and minimum possible. For each *ex-ante* portfolio selection model the expected returns are based on the weighted means for each security, so that portfolio composition varies only because of the differences in the representation of expected dispersion as measured by the slope parameter of the characteristic line.

The return of each portfolio generated by each selection model was calculated for each year of the *ex post* period. The mean and standard deviation of these returns for each portfolio were calculated, then for each *ex*

[1] Another sample of 150 securities, including the original 75, was also tested, but the upper bound constraint applied in the original research seems to have caused superfluous diversification [2], and so these results are not reported.

TABLE 1. *Ex post* STANDARD DEVIATION—RETURN REGRESSION

Model	Interdcept (*t*-statistics in Parentheses)	Slope	Standard error of estimate (adjusted for degrees of freedom)	R^2 (adjusted for degrees of freedom)
Covariance	1·729 (1·17)*	0·970 (12·30)	0·630	0·852
Index Quadratic	4·585 (2·57)	0·829 (8·91)	0·547	0·744
Index Linear-OLS	12·746 (6·46)	0·433 (4·21)	0·685	0·391
Index Linear-MSAE	1·951 (1·86)	0·940 (15·84)	0·431	0·906

* Not significantly different from zero at *P*-level < 0·95 in a two tailed test.

ante efficient set, the *ex post* standard deviations were regressed on the *ex post* means [1, pp. 186–187], with the results shown in Table 1. Note that the coefficient estimates for the Covariance and Index Linear—MSAE models are very similar, while the coefficients of Index-Linear OLS are very different. It would be desirable to test whether these regression lines are coincident, but the standard error of estimate of the second model implies a variance that is significantly smaller than each of the other models so that data cannot be pooled for analysis of variance. In fact, a variance ratio test of the estimated variances pairwise rejects the null hypothesis that the variances around the regression lines are the same, generally at the P-level $> 0{\cdot}95$. Because of the limited sample, only tentative conclusions may be drawn: first, that MSAE estimation appears clearly to dominate the OLS estimation as a method of *ex ante* efficient portfolio construction; second, a linear model using MSAE estimates may be as effective as the full covariance representation, but very much less costly to compute.

ACKNOWLEDGEMENTS

The author is indebted to Professors KJ Cohen and JA Pogue for permission to use their data for this study, and to use some of their results for comparative purposes.

REFERENCES

1. Cohen KJ and Pogue JA (1967) An Empirical Evaluation of Alternative Portfolio-Selection Models. *J. Bus.* **40** 166–193.
2. Evans JL and Archer SH (1968) Diversification and the Reduction of Dispersion: An Empirical Analysis. *J. Finance.* **23** 761–769.
3. Fama EF (1965) Portfolio Analysis in a Stable Paretian Market. *Mgmt Sci.* **11** 404–419.
4. Fama EF (1965) The Behavior of Stock Prices. *J. Bus.* **38** 34–105.
5. Sharpe WF (1963) A Simplified Model for Portfolio Analysis. *Mgmt Sci.* **9** 277–293.
6. Sharpe WF (1967) A Linear Programming Algorithm for Mutual Fund Portfolio Selection. *Mgmt Sci.* **13** 499–510.
7. Sharpe WF (1971) Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios. *Mgmt Sci.* **18** B-1–B-13.

JC Wiginton
(*October 1975*)

School of Business
Queen's University
Kingston
Canada K7L 3N6

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

ARNOLD V (1972–73) *German: Ein Beitrag zur Theorie der Externen Effekte.* PhD Thesis, Georg-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 3400 Göttingen, Prinzenstrasse 1, Germany)

CHAPMAN JB (1974) A Comparative Analysis of Male and Female Leadership Styles in Similar Work Environments. Doctoral Thesis, University of Nebraska. (Available at: the Love Library, University of Nebraska, Lincoln, NE 68508, U.S.A.)

CLARK RF (1972) An Economic Evaluation of Zoning as a Land-Use Policy Tool in the Rural–Urban Fringe. MSc Thesis, Ontario Agricultural College. (Available at: the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

DICK E (1970–72) *German: Untersuchungen einiger Grundproblems der Wohlfahrtsökonomik.* PhD Thesis, Georg-August-Universität Gottingen. (Available at: Nedersächsischen Staats- und Universitäts-Bibliothek, 3400 Göttingen, Prinzenstrasse 1, Germany)

FISCHER V (1970–72) *German: Die Problematik der Behandlung von Second-Best-Problemen mit Hilfe von Optimierungsmodellen.* PhD Thesis, Georg-August-Universität Göttingen. (Available at: Niedersächsischen Staats und Universitäts-Bibliothek, 3400 Göttingen, Prinzenstrasse 1, Germany)

HESSE K (1972–73) *German: Das Diskretionäire Einkommen, seine Bestummung und Verwendung.* PhD Thesis, Georg-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 3400 Göttingen, Prinzenstrasse 1, Germany)

JOB MO (1973) A Linear Programming Model for Evaluating the Capital Requirements of Small Crop Farms in Trinidad. MSc Thesis, Ontario Agricultural College. (Available at: the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

LAVOIE R (1974) *French: Crédibilité de la Source, Disparité et Efficzcité de la Communication.* MSc Thesis, Université de Sherbrooke. (Available at: the General Library, Université de Sherbrooke, Québec J1K 2R1, Canada)

LUPIEN S (1973) *French: L'Influence de la Personnalité et de l'Image de la Marque sur le Choix de la Bière.* MSc Thesis, Université de Sherbrooke. (Available at: the General Library, Université de Sherbrooke, Québec J1K 2R1, Canada)

MICHON C (1974) *French: L'Image Iconique dans le Processus de Choix d'une Marque Nouvelle.* MSc Thesis, Université de Sherbrooke. (Available at: the General Library, Université de Sherbrooke, Québec J1K 2R1, Canada)

REDINBAUGH LD (1970) Changes in Marketing Strategies and Tactics of Life Insurance Companies. Doctoral Thesis, University of Nebraska, (Available at: the Love Library, University of Nebraska, Lincoln, NE 68508, U.S.A.)

ROMAGNAN JM (1973) *French: Application de l'Approche Marketing à une Institution Universitaire: Etude des Relations de Service de l'Université de Sherbrooke.* MSc Thesis, Université de Sherbrooke. (Available at: the General Library, Université de Sherbrooke, Québec, J1K 2R1, Canada)

SAVU ATP (1972) Adaptive Intertemporal Equilibrium Models for Policy Planning and Review. MSc Thesis, Ontario Agricultural College. (Available at: the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

SMITH RL (1970) The Process of Structural Change in Manufacturing Organizations. Doctoral Thesis, University of Nebraska. (Available at: the Love Library, University of Nebraska, Lincoln, NE 68508, U.S.A.)

TALLENT DR (1970) The Effect of Decentralization on the Managerial Functions of Planning, Direction, and Controlling. Doctoral Thesis, University of Nebraska. (Available at: the Love Library, University of Nebraska, Lincoln, NE 68508, U.S.A.)

WEST JS (1972) Marketing U.S. Goods to Eastern Europe: The Attitudes of U.S. Businessmen. Doctoral Thesis, University of Nebraska. (Available at: the Love Library, University of Nebraska, Lincoln, NE 68508, U.S.A.)

The following are obtainable at the Yvette Jacobson Library, Management Centre, University of Bradford, Emm Lane, Bradford, Yorkshire BD9 4JL, U.K.:

BALLANGER NJ (1973) U.K. Financial Institutions and the Management of U.K. Industry. MSc Thesis, University of Bradford.

CARNALL CA (1973) Mechanized Assembly in the Metalworking Industries. MSc Thesis, University of Bradford.

DARWIN DB (1974) A Study of Money and Job Enrichment as Sources of Motivation to Work. MSc Thesis, University of Bradford.

ELLSMORE MA (1974) Evaluating Logistics Service—A Model Building Approach. MSc Thesis, University of Bradford.

FARRER PJ (1972) Productivity Bargaining—A Feasibility Study. MSc Thesis, University of Bradford.

FORD D (1971) The Marketing of Central Heating with Particular Reference to the Channels of Distribution. MSc Thesis, University of Bradford.

HALSTEAD AS (1971) Exporting and the Multinational Company: a Study of the Institute of Export. MSc Thesis, University of Bradford.

JEFFERIES JA (1971) The Role of the Retailer Sponsored Label in the Marketing of Domestic Appliances. MSc Thesis, University of Bradford.

LOCKYER DE (1971) Shopping Hierarchies and Retail Location Decisions. MSc Thesis, University of Bradford.

MARGERISON CJ (1971) The Development and Evaluation of an Experimental Social Studies Programme in Primary Education. PhD Thesis, University of Bradford.

TEMPERLEY J (1972) The Application of Computers to Statistical Sales Forecasting with Reference to the Engineering Industry. MSc Thesis, University of Bradford.

WHITELAW M (1971) The Methodology of the Evaluation of Management Training. MSc Thesis, University of Bradford.

The following are obtainable at the University of California Library, Berkeley, California 94720, U.S.A.:

ABED MI (1972) The Stochastic Cash Balance Problem. PhD Dissertation, University of California, Berkeley.

ASSMUS G (1971) An Adaptive Media Model. PhD Dissertation, University of California, Berkeley.

DESHMUKH SD (1971) Computations in Stochastic Environments: An Investigation of Sequential Decision-Making Under Uncertainty. PhD Dissertation, University of California, Berkeley.

FRIEDMAN I (1971) Application of Linear Programming to Investment Problems with Non-linear Properties. PhD Dissertation, University of California, Berkeley.

HALAL WE (1971) The Compatibility of Various Forms of Leadership, Technology, and Motivation. PhD Dissertation, University of California, Berkeley.

HENDERSON JY (1972) Conceptual Problems of a National Balance Sheet for Australia. PhD Dissertation, University of California, Berkeley.

JEWETT JE (1972) College Admissions Planning: Use of a Student Segmentation Model. PhD Dissertation, University of California, Berkeley.

MCKIBBEN WL (1971) Econometric Forecasting of Fundamental Operating Data and Their Function in Stock Market Price Formation. PhD Dissertation, University of California, Berkeley.

SIMS JH (1971) Differentiating Characteristics of High-Performing and Low-Performing Engineers. PhD Dissertation, University of California, Berkeley.

SWANSON EB (1971) On Being Informed by a Computer-Based Management Information System: A Study in Involvement and Appreciation. PhD Dissertation, University of California Berkeley.

WAAGE F (1971) Transient and Steady State Decision Rules of a Bayesian Consumer Who Buys in Binomial Markets. PhD Dissertation, University of California, Berkeley.

WALKER JM (1970) Models of Academic Governance in Emergent Universities. PhD Dissertation, University of California, Berkeley.

The following are obtainable at the Department of Management Sciences Library, University of Manchester Institute of Science and Technology, PO Box 88 Manchester M60 1QD, U.K.:

BIBBY EL (1973) The Role of the Professional Engineer. MSc Thesis, University of Manchester.

BRENES E (1973) Analysis and Critique of Difference Approaches to Human Resource Accounting. MSc Thesis, University of Manchester.

GARCIA-TORRES IA (1973) The Transferability of Technology and Management to Underdeveloped Countries as Elements of Economic Development. MSc Thesis, University of Manchester.

INTZOGLOU A (1973) Transfer Pricing—A Survey of The Literature. MSc Thesis, University of Manchester.

MARCHINGTON MP (1973) Some Determinants of Work Group Effectiveness. MSc Thesis University of Manchester.

NESSIM EDM (1973) A Study of the Cost of Capital under Conditions of Risk. MSc Thesis, University of Manchester.

PRICE BW (1973) The Development of an Analytical Technique as a Guide to Management Policy for a Multi Item Macro Inventory. MSc Thesis, University of Manchester.

SHIPPEY JA (1973) The Effects of Interaction in the Learning Situation. MSc Thesis, University of Manchester.

SYMEONIDES S (1973) Operational and Cost Analysis of Coronary Care Service. MSc Thesis, University of Manchester.

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 30 Memorial Drive, Cambridge, MA 02139, U.S.A.:

ANDERSSEN AH (1972) Preliminary Design of a Human Resource Accounting System for the United States Navy. SM Thesis, MIT, Cambridge, MA.

ANDO S (1973) Joint Exchange-Rate Management by European Community Central Banks. SM Thesis, MIT, Cambridge, MA.

BAILLY JP (1970) Computer Models in Strategic Planning: A Functional Analysis. SM Thesis, MIT, Cambridge, MA.

BARNETT TO (1970) The Relation of Organizational Relationships to Computer Planning Effectiveness. SM Thesis, MIT, Cambridge, MA.

DAVIDSON LM (1973) The Process of Employing the Disadvantaged. PhD Thesis, MIT, Cambridge, MA.

FROHMAN AL (1970) The Development and Impact of a Joint Goal Setting Technique for Conflict Resolution. PhD Thesis, MIT, Cambridge, MA.

GERRITY TP (1970) The Design of Man–Machine Decision Systems. PhD Thesis, MIT, Cambridge, MA.

IACUELLI D (1970) Management Factors in Joint Ventures: Italy, A Case Study. PhD Thesis, MIT, Cambridge, MA.

LUCAS HC (1970) The Integration of Data Processing and the Organization. PhD Thesis, MIT, Cambridge, MA.

MAISEL ME (1970) Examination of the Motivational Basis of Performance. SM Thesis, MIT, Cambridge, MA.

MAY RC (1970) Simulation Studies Toward Production Control in Complex Manufacturing Facilities. SM Thesis, MIT, Cambridge, MA.

NAGORNIAK JJ (1970) The Application of Sharpe's Capital Asset Pricing Model to Debt/Equity Markets. SM Thesis, MIT Cambridge, MA.

OWENS RC (1971) Primary Access Control in Large-Scale Time-Shared Decision Systems. SM Thesis, MIT, Cambridge, MA.

PERKINS AE (1971) Organizational Environment and Design Considerations for Management Information Systems Development. SM Thesis, MIT, Cambridge, MA.

ROSEMAN LD (1972) Strategic Change in Financial Intermediaries. SM Thesis, MIT, Cambridge, MA.

ROSENBERG JL (1972) Evaluation of an Automatic Programming System for Management Information Systems. SM Thesis, MIT, Cambridge, MA.

SUTHERLAND RL (1973) An Analysis of Business Planning of New Technology-Based Enterprises. SM Thesis, MIT, Cambridge, MA.

THOMPSON KA (1973) An Environmental Analysis of the International Pharmaceutical Industry. SM Thesis, MIT, Cambridge, MA.

The following are obtainable at the Lockwood Library, State University of New York at Buffalo, 3435 Main Street, Buffalo, New York 14214, U.S.A.:

ALBRIGHT P (1970) The Impact of Unions on Wages of Supermarket Cashiers in Southern Ontario. PhD Thesis, University of New York at Buffalo.

ARBEIT S (1971) The Occupational Identity of the Retail Pharmacist: An Inquiry into Its Nature, Correlates and Consequences. PhD Thesis, University of New York at Buffalo.

GODIN W (1972) Benefits to Management Through Improved Cost Accounting Information: An Analysis and Case Study. PhD Thesis, University of New York at Buffalo.

GOMOLKA E (1974) The Effect of Personality Systems of Small Business Company Presidents as a Determinant of Organization Structure. PhD Thesis, University of New York at Buffalo.

GREENE F (1974) An Analysis of the Functioning of the Fact Finding Process in Selected Schoolboard–Teacher Negotiation Impasse in New York State. PhD Thesis, University of New York at Buffalo.

HIERL A (1974) The Effects of Warrant Financing of the Valuation Firm. PhD Thesis, University of New York at Buffalo.

HREBINIAK L (1971) A Multivariate Analysis of Professional and Organizational Commitment Orientations Among Teachers and Nurses. PhD Thesis, University of New York at Buffalo.

JATUSIPITAK S (1972) Stock Price Behavior: An Empirical Test on the Non-Stationery Log-normality Hypothesis. PhD Thesis, University of New York at Buffalo.

JOHNSON R (1970) The Extinguishment of Convertible Bonds. PhD Thesis, University of New York at Buffalo.

KENDALL K (1974) A Systems Study of Blood Inventory Management. PhD Thesis, University of New York at Buffalo.

PERCIVAL J (1973) Intertemporal Analysis of Corporate Bond Risk Premium. PhD Thesis, University of New York at Buffalo.

PALMER D (1974) Teachers, Researchers and Consultants: An Empirical Study of the Faculty in Collegiate Schools of Business. PhD Thesis, University of New York at Buffalo.

RAJARAMAN AS (1972) Measurement of Cost of Uncertainty in Stochastic Dynamic Programming Modes. PhD Thesis, University of New York at Buffalo.

ROSE J (1971) Organizational Union Representation Elections. PhD Thesis, University of New York at Buffalo.

SHICK RA (1972) Evaluating the Financial Aspects of Mergers and Acquisitions: An Application of the Principles of Valuation. PhD Thesis, University of New York at Buffalo.

SMITH C (1970) Demand for Short-term Government Bonds. PhD Thesis, University of New York at Buffalo.

TUROFF S (1970) Growth of Manufacturing in Connecticut: 1958–63. PhD Thesis, University of New York at Buffalo.

WHIPPLE T (1971) A Multi-dimensional Analysis of Home Buying Decision Making. PhD Thesis, University of New York at Buffalo.

YERKES R (1974) Variations in Adherence to Expressed Rules of Ethics by the Accounting Profession and Some Factors Associated with Such Variation. PhD Thesis, University of New York at Buffalo.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 1, 1976. Pergamon Press. Printed in Great Britain

ANNOUNCEMENT

IX INTERNATIONAL SYMPOSIUM ON MATHEMATICAL PROGRAMMING

Budapest—23–27 *August* 1976

General information

The IX International Symposium on Mathematical Programming will be held 23–27 August 1976, at the Hungarian Academy of Sciences, Budapest, Hungary. This International Symposium will continue a series of research conferences begun in 1949 in Chicago and held last in 1973 in Stanford.

The Symposium will be devoted to the theory and applications of mathematical programming including linear programming, nonlinear programming, integer programming, stochastic programming, dynamic programming, optimization in networks, related problem in graph theory and combinatorics, game theory, theory of convex sets and convex polytopes, mathematical programming software, applications in economics, environmental systems, engineering, biology, etc.

The official languages are English, French, German and Russian. Translation will be provided from French, German and Russian to English. In order to reduce translation to the minimum, speakers of the Symposium are kindly requested to use English.

For further information write to:

Dr. G Mitra
Department of Statistics & OR
Brunel University
Uxbridge
Middlesex UB8 3PH

or

Dr. HP Williams
Department of Statistics & OR
University of Sussex
Falmer
Brighton, Sussex

or IX International Symposium on Mathematical Programming
Bolyai János Mathematical Society
H 1368 Budapest, POB 240
Hungary

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

EDITORIAL

In-Company Training

UNFORTUNATELY, I don't have reliable statistics for the expenditures of industrial companies on training, as it would have been instructive to make sensible comparisons between different enterprises or industries, not only with respect to training expenditure per employee or the percentage of the total budget devoted to training, but also in order to establish whether trends are at all discernible over a period of time and whether some enterprises or industries are becoming more training-conscious than others, and why. Such evidence as there is suggests that many large companies take training very seriously and are prepared to devote substantial resources to it. Some have permanent facilities allocated to this activity, in the form of training schools or staff colleges with elaborate physical amenities, including living accommodation for trainees, full-time teaching and non-teaching staff, and even suitable recreational fixtures which are regarded both as suitably diversionary as well as an essential part of the training programme.

There are two questions one is tempted to ask: first, why do these companies believe it preferable to build up their own elaborate facilities instead of sending their employees to external courses run by various consultants and educational agencies, and secondly what benefits do they precisely hope to derive from their in-company training?

The two questions are not mutually exclusive and it is impossible to discuss the first without the second. But let us assume that a systematic examination in a company has led to the identification of the needs of a particular individual; it should then be a simple matter to compare the relative merits of external with internal courses in the light of such needs and to determine which appears to offer the best fit in terms of admission requirements, content and length, the level and manner of treatment of the subject matter, the expected level of knowledge or skill to be acquired, the other likely participants and the ex-curricula benefits that might accrue. In some cases the choice is clear cut: the particular expertise or novelty offered by some courses, or their image and status, may be such that they override other considerations, and even when there is a glaring disparity between the corresponding costs, this is usually not a dominant factor in the final decision. But when external courses do not possess

any unique features that distinctly put them a cut above internal courses, the latter are almost invariably preferred by executives in the personnel function responsible for training policy.

One reason for this is that where internal training programmes exist, it is natural to regard them as a fixed-cost facility, which any cost-effectiveness conscious management would want to utilize to the full. The internal course then becomes like any other internal product, which—as long as it is not significantly inferior to a competing external product—should be the natural choice of all parts of the enterprise in demand for such a product. A second (and often regarded as a more powerful reason) is that the internal course is thought to be better equipped to cater for the needs of the organization and its members, because it is designed by internal people, who are familiar with the objectives of the enterprise and have a good appreciation of its problems. Thus, unlike the external course which is aimed at participants from many firms, even from different industries, with very diverse backgrounds and experiences, the internal course can be better structured around common interests, be unhampered by problems of confidentiality, and—as an internal facility—be more flexible to respond to specific needs of individuals.

On the other hand, there are many advantages which external courses have and which merit due consideration. The fact that a choice has to be made demonstrates the need to delineate the precise reasons for the enterprise seeking to develop training programmes for its employees and to have an understanding of the multi-objectives that underline such schemes.

Vol. 4
1976

A useful starting point is the widely held proposition that the main task of the personnel function is to compare the attributes possessed by people in the organization with the range of attributes that are required for its current and future well-being and then to take appropriate action to fill the gaps. On a global company scale, and bearing in mind future plans, such an exercise can help to formulate a recruiting policy, but the matching matrix of existing versus desirable attributes can also be constructed for each member of the organization as a basis for career-path planning. The required attributes fall into four categories:

- technical skill, knowledge of a given job or an array of jobs and the way they interact with other jobs;
- experience in exercising authority and responsibility in certain jobs and in a given environment;
- personality, leadership, decisiveness, social skills in communication;
- certain attitudes regarding the ethos of the enterprise and its mission.

It is very unlikely that the matching matrix for any given individual remains static for long. With the passage of time and the effect of normal progression in the organizational hierarchy, the individual faces new prospects, and the range of attributes he is expected to possess must change with each new assignment or

promotion. Furthermore, the enterprise as a whole does not remain static and new demands may be placed on its reorganization as the internal and external environments change. This dual process of adaptation, of the organization to adjust its capabilities to face new challenges and of the individual to learn whatever new roles he would need to play in the organizational structure, inevitably reveals deficiencies in the matching matrix, which in some cases can be adequately cured only by accelerated employee turnover, but in the main the organization resorts to designing training programmes as a means of improving employees' aptitude to perform current and future tasks. Thus, we can identify ten possible uses of training:

1. Imparting substantive knowledge on techniques and new developments (for employees who need to learn particular skills or to implement new methods, and for those who do not need detailed technical knowledge on a given topic but would benefit from appreciation courses);
2. Transmitting information about the enterprise and its constituent parts, about organizational changes, new procedures, future plans;
3. Identifying problems of which people may be unaware, because of organizational barriers or lack of communication;
4. Evaluating information, in terms of its accuracy, timeliness, validity and possible use (the relationship of information to the decision process is particularly valuable as an exercise for trainees, when they are set a problem and they have to specify what information would be needed to solve it or to control a given operation);
5. Problem solving, either to demonstrate the applicability of certain methodologies in relation to particular problems, or to improve the capability of executives in problem solving in general (case studies taken from the enterprise's own history or formulated within the current operational context are often regarded as effective means of teaching problem solving techniques);
6. Providing 'group therapy' sessions for people to 'let off steam', to discuss their problems in front of others and get explicit or implicit feedback from fellow participants and from course tutors, particularly on behavioural aspects;
7. Developing social skills, on how to communicate with other people, how to present information and improve one's techniques in public speaking, how to participate in an interview (as an interviewer or as an interviewee), and how to negotiate;
8. Appreciating teamwork and conflict resolution, making people aware that they have to work together (for example by presenting them as a

group with a problem, for which they need to organize the group, allocate tasks and responsibilities to individuals within it and cope with conflicts that emerge from the way in which the group is organized or from the definition of the problem);

9. Promoting *esprit-de-corps*, a feeling of belonging to the enterprise and a sense of pride in its progress and status in the industrial scene;
10. Organizing brainstorming sessions to deal with unresolved problems, to throw a new light on possible new opportunities, to proceed sometimes through an unstructured discussion to a sudden realization of how problems can be redefined and handled.

Put in this way, we begin to appreciate that people should not be sent on a course just because 'it would be good for them', but that in each case the precise purpose should be properly thought out, in order to increase the probability that individuals, as well as the organization as a whole, can reap the potential benefits that a course is designed to provide. Alas, it is often difficult to ascertain the extent to which courses are successful in achieving their purpose. Admittedly, when a course is aimed at imparting substantive knowledge or improving certain skills, it is possible to test the course participants before and immediately after the course and establish (within the limitations that such tests are subject to) how much the participants have learnt from the course; it is even possible to test them again several months later to determine how much knowledge is retained and perhaps to take action to reduce the rate of attrition. Similarly, when the course is aimed at changing attitudes, it is possible through questionnaires and a series of interviews to ascertain whether the desired change has materialized, although the measurement methodology may not always be as rigorous or as objective as one would like. And with respect to many of the other purposes listed earlier, the rigour in measuring whether a course has achieved its stated objectives in relation to a particular individual or group of individuals is bound to be questionable. We have to admit that in reality personnel managers have only a hazy notion regarding the reasons for sending an employee on a course and an even hazier notion regarding its effects. This is partly due to the reluctance of personnel managers to be too explicit about their objectives, and partly due to the inherent difficulty of devising a convincing cost–benefit analysis in each case.

In the light of this discussion, and the array of purposes listed above, it is understandable why many firms prefer in-company training to sending employees on external courses. Not only are some of these stated purposes so intimately associated with the internal environment that external agents would be ill-equipped to provide the necessary link between the theme of the course and current internal problems, but it is argued that assessment of course achievement is bound to be more reliably obtained for internal courses (irre-

Vol. 4
1976

spective of how crude the assessment is) and hence future courses can be more easily modified or restructured. For many purposes such logic is difficult to refute and internal courses score every time compared with external alternatives. The danger is that, taken to the extreme, in-company training can become a breeding ground for inward looking, dogmatic and insulated organization-men, so engrossed with their internal problems that they may miss the stimulus which external agents can successfully provide. There is clearly a need for a mixture of internal and external facilities, and—in spite of the difficulties alluded to—for a continuous reappraisal of objectives and achievements of training programmes.

SAMUEL EILON
Chief Editor

ol. 4
976

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

'Irrationality' of Managerial Judgments: Implications for Information Systems[1]

HERBERT MOSKOWITZ[2]

RALF E SCHAEFER[3]

KATRIN BORCHERDING[3]

(*Received August 1975; in revised form November 1975*)

Four principal factors affecting rational information processing behavior in organizations are discussed. These are: (1) the nature and attributes of the messages received from the Management Information and Decision System (MIDS); (2) The personal qualities and attitudes of the manager himself; (3) the degree of social interaction; and (4) if a collective is involved, the format and group decision function used in arriving at a group position. Ways in which information specialists and operations researchers can counter, ameliorate, and exploit managers' judgmental biases and inconsistencies are then proposed. Areas for further research are suggested.

AN AREA of neglect and ever increasing concern in the design of management information and decision systems (MIDS) is the psychology of the manager *qua* judge and decision maker. This consideration has become specially relevant and important to operations researchers and management scientists as they increasingly turn their energies toward building MIDSs for solving the 'wicked', unstructured problems in organizations [11], [25]. A key characteristic of such problems is uncertainty. Most of the models developed for these problems require subjective inputs, often expressed as probabilities or probability distributions. The *sine qua non* of such models' value rests on the quality of these subjective inputs. Good probability assessments become especially critical when the output from one model in the form of a prediction or decision serves as input data for higher level and/or other organizational decisions. Here errors can become dangerously amplified throughout the organizational hierarchy

[1] This is a slightly abridged version of the original paper. The unabridged version, which contains a comprehensive list of references, can be obtained by writing to Herbert Moskowitz at Purdue.

[2] Krannert School of Industrial Administration, Purdue University, U.S.A.; Guest Professor at the Sonderforschungsbereich (SFB) 24 'Sozial- und Wirtschaftspsychologische Entscheidungsforschung', at the University of Mannheim.

[3] Sonderforschungsbereich (SFB) 24 'Sozial- und Wirtschaftspsychologische Entscheidungsforschung', at the University of Mannheim, Germany.

[5]. A familiar example of the state of affairs that poor subjective inputs can create is seen in research and development (R&D). In most cases, project management models like PERT have not prevented or even mitigated the large cost and schedule overruns usually experienced. Nor have any of the many stochastic R&D decision models been implemented in practice, either due to the lack of a historical data base or 'poor' subjective inputs [3].

Design shortcomings of MIDS

Recognizing the pivotal role that managers play as integrators or processors of information for decision making, modern technology has been mobilized to help increase their ability to judge and decide by supplying them with a proliferation of data in the most timely and efficient manner. The underlying premise has been, "The *more*, *quicker*, and *higher quality* the information, the better the judgments and decisions." Manifestations of these efforts in organizations are seen in the growth of computerized MIDSs and the pervasive use of groups and experts. Unfortunately, instances abound where the provision of more timely, complete, and reliable information has not improved decision making. There are at least three basic reaons for this:

1. MIDS designers are ignorant of the needs and capabilities of the manager, the recipient and user of the information [1].
2. Operations research models are all too frequently developed with no consideration given to information system requirements and capabilities (i.e. the information and decision systems are treated independently rather than as a unified whole) [11]. Nor is any consideration really given to the behavioral characteristics and capabilities of the managers who provide the subjective inputs and who perform as an integral component of the MIDS [25].
3. Management control systems in organizations generally discourage rather than encourage accurate assessment of uncertain events. For example, Woods [28] and Souder [21] found that firms' control systems produced systematic biases in the form of underestimation and undue optimism.

Vol. 4
1976

Need for behavioral research

The foregoing underscores the need to understand better, both *psychologically* and *sociologically*, the relationship between man and information if MIDSs are to be improved. We must thus ask, "What are man's limitations as an information processor and how is his behavior affected by the informational environment and social situation in which he functions?" Surprisingly, much of this area has received little attention and has been virtually ignored in MIDS design.

One purpose of this paper is to summarize some effects of various informational characteristics, social factors, and personal attributes of the manager on human information processing behavior and to discuss their implications with

respect to MIDS design. Many of the results derive from a series of studies which examined human information processing behavior in a simulated financial MIDS environment. The studies were also designed, in part, to determine whether certain behavior observed in the psychological laboratory was representative of behavior in the 'real world'.

A second purpose is to suggest several options open to MIDS designers and operations researchers to improve MIDS's by incorporating, countervailing, attenuating, and exploiting human foibles and limitations and tendencies to subvert such systems.

RESEARCH PERSPECTIVE

A MIDS can be broadly defined in terms of the following principal components (Fig. 1): (1) An *information source*—the organization for which decisions are to be made and from which information is collected. (2) *Data*—the process of observing, measuring, and recording information about the source. (3) *Prediction*—the process of estimating and drawing inferences. (4) *Values*—the evaluation of these predictions with regard to the objectives or goals of the organizations. (5) *Action*—the choice of a course of action. (6) *Outcome*—the consequences of the action which was taken. The major focus of the research concerned the third MIDS component—prediction. However, the implications of the results cannot be decoupled from the other MIDS components.

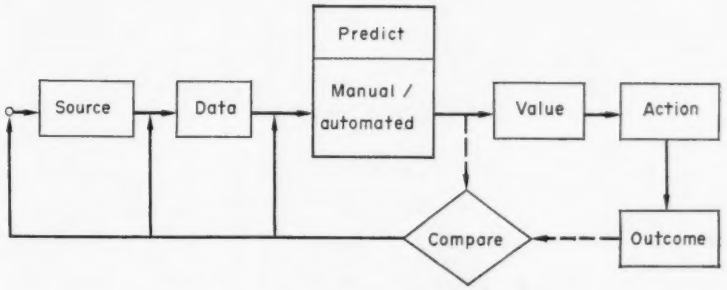


FIG. 1. *Management information system (MIDS).*

A model for organizing research on the prediction component is depicted in Fig. 2. Briefly, a message characterized by a vector of attributes (such as informativeness, complexity, timeliness, etc.) is received by an information processor from an information source. This message along with the processor's accumulated knowledge is then processed into a judgment. Additional inputs impinging on the information processor include his inherent motivations, biases, perceptions, and cognitive limitations as well as external social interaction

forces (e.g. isolation or presence of others, feedback or no feedback on previous judgments, discussion or no discussion of judgment to be made with others, etc.). All these factors bear on *how* and *how well* the messages received are processed into a judgment. In the case of groups, a group decision function for amalgamating individual member beliefs into a common position is also involved in arriving at the ultimate judgment. Moreover, in both individual and group circumstances, the correspondence between the cognitive judgment and the articulated judgment depends on the reward function. Unfortunately, there is no way of knowing whether the verbalization and the belief are isomorphic. Through the use of a proper reward structure (e.g. scoring rules [26]), however, one can hopefully achieve isomorphism.

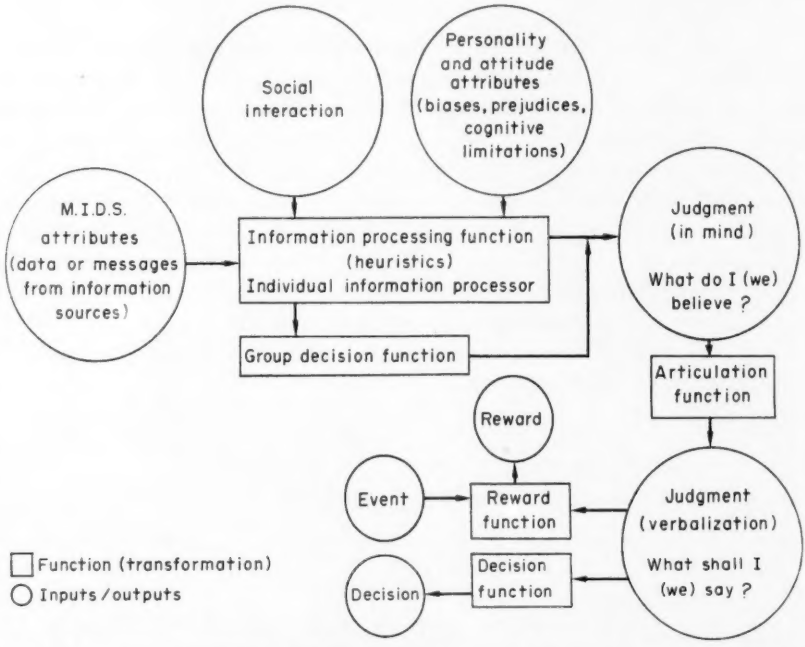


FIG. 2. *Some motivational and cognitive factors impinging on information processing behavior.*

BEHAVIORAL LITERATURE

Although there was little research in the area of human information processing prior to 1960, since then much has been done, and the annual volume of research has been increasing exponentially, stimulated by the growing awareness of the problem's significance and the aid of the omnipresent computer. An excellent summary of this research up to 1971 is found in Slovic and Lichtenstein [20]. Much of the recent work has been accomplished within the 'Bayesian' school by

Vol. 4
1976

psychologists, where Bayes' Theorem has served as the baseline paradigm for opinion revision upon receipt of new information. The modern impetus for what we are calling the Bayesian paradigm is rooted in the work of von Neumann and Morgenstern [24] who revived interest in maximization of expected utility as a core tenet of rational decision making, and to Savage [15], whose *Foundations of Statistics*, fused the concepts of utility and subjective probability into an axiomatic theory of decision making under uncertainty. One aspect of this rapidly emerging school of thought called 'Decision Theory' asserts that we ought to cast our opinions (beliefs) about the world in probabilistic terms. When we translate our opinions into probabilities, a mathematical formula, Bayes' Theorem, dictates the optimal way that our estimates should change upon receipt of new information.

This line of inquiry has turned up a number of consistent tendencies associated with human information processing behavior. One such tendency is described by the notion of *conservatism*. Conservatism denotes that upon receipt of new information, humans revise their beliefs in the same direction as Bayes' Theorem prescribes, but that the revision is typically too small. Much of the Bayesian research has been motivated by a desire to discover the determinants of conservatism.

Unfortunately, the work in this area, until recently, has been largely confined to the psychological laboratory where introductory psychology students served as subjects, and the tasks employed have had little relevance to real world decision problems. For example, most Bayesian experiments have been based on statistical data-generating paradigms employing such devices as book bags and poker chips, etc. [20]. Whether other groups of people behave similarly in more realistic decision settings remains an important empirical question, which is now receiving increasing attention [20], [27]. Moreover, little systematic attention has been paid to the effect of such organizational factors as the nature and characteristics of the information system, social environment, and social interaction on information processing behavior.

As a result, there is a pressing need to replicate and extend the various types of studies performed in the laboratory to specific organizational settings. If the results of such research confirm the findings in the psychological laboratory, then such persistent tendencies may have sufficient generality to qualify as propositions or laws. These could then serve as guidelines for integrating behavioral factors in MIDS design and for training managers as information processors or to provide reasonable inputs for information processing by a formal model. Such studies would also provide further insights into the nature of man's cognitive processes.

MIDS ENVIRONMENT

The decision context in our investigation, from which much of the evidence in this paper is drawn, involved the evaluation of applicants for personal loans

by a bank lending officer. One of the fundamental judgments that a loan officer (or loan committee) must make in processing loan applications is an assessment of the likelihood that an applicant will either default or be delinquent in repaying his loan. Lending officers arrive at this judgment on the basis of the data that they receive about the applicant. The loan officer has several different sources upon which to base his judgment. For this particular task setting four of these sources were used: (1) the prospective borrower's application form, (2) the bank's internal records, (3) a credit scoring system based on the borrower's attributes, and (4) an independent retail credit evaluation. Reports from each of these sources were provided in the form of objective (relative frequency) conditional probabilities relating the possible messages to past delinquencies and non-delinquencies of previous borrowers. These sources and messages were basically the same as those available to and used by many bank lending officers.

In the experiments, each subject acting as a bank loan officer or loan committee member assessed the probability that a loan applicant requesting a $5000, unsecured, vacation loan would become delinquent in his payments during the coming year. His assessments were based on the background information obtained from the loan application and data on the applicant from the three conditionally independent, binary information sources mentioned. Illustrations of the information sources are shown in Table 1. Further details on the experimental tasks and the development of the normative Bayesian probability revision model are found in [10] and [12].

TABLE 1.

Data / Hypothesis	Credit data service (retail information)		Bank's internal records		Borrower's attributes	
	X	X′	Y	Y′	Z	Z′
H (delinquent)	0.20	0.80	0.10	0.90	0.30	0.70
H′ (not delinquent)	0.80	0.20	0.90	0.10	0.70	0.30

X Good rating
X′ Bad rating
Y Never delinquent in past 5 years on loans, etc.
Y′ Delinquent at least once on loans, etc.
Z Score of greater than 75 points
Z′ Score of equal to or less than 75 points

OVERVIEW OF RESULTS

The findings of these bank lending studies, to be briefly summarized and contrasted to existing literature, concern the effects on information processing behavior of: (1) the message attributes of the MIDS information sources, (2) an individual's characteristics and attitudes, (3) the social environment and degree

of social interaction, and when groups are involved (4) the group decision function for aggregating individual member judgments into a group judgment. The subjects used in these studies varied, but mainly included upper level undergraduate business and engineering students, and graduate industrial management students. Most had some prior work experience; some had actually worked in financial institutions and were former bank loan officers. All groups were limited to three members.

Before discussing the effects and implications of the above factors it is worth pointing out that the results from these studies in general confirmed the conservatism findings in the psychological laboratory for both individuals and groups [10], [12]. This phenomenon thus appears to have considerable generality. It is of further interest to note that interacting consensus groups (a common group format), were found to be more conservative than individuals or artifactual groups (randomly selected individuals whose responses were averaged to obtain a group response). This finding, too, has been supported in the psychological laboratory [6].

Message attributes

The MIDS designer has considerable control over the messages sent to a decision maker, and thus can influence as well as compensate for their psychological effects through his design. To do this, he must understand the relationship between the attributes of his system and human biases in processing the messages received into judgments.

1. *Informativeness.* As found in many psychological experiments [20, pp. 701–703], conservatism increased with increased message informativeness with individuals as well as groups [12]. The numerous replications of this finding would seem to qualify it as a generally consistent human tendency.

2. *Order effects.* As in most psychological studies, individuals exhibited primacy effects, i.e. gave earlier received messages more salience than later received messages [10]. Moreover, the order in which both positive (confirming evidence with respect to the hypothesis in question) and negative information (disconfirming evidence) was received, influenced the magnitude of this effect. Similar order effects were found with artifactual groups, but order effects were vitiated with interacting consensus groups.

While it is clear that the sequence in which messages are received can influence individual beliefs, the nature of its effect is less clear, as recency effects (later received messages are given more salience) have also been observed in several experiments [20, p. 704]. It is even more difficult, however, to generalize on sequence effects for groups because of the dearth of such experimentation. It has been found that the group format and group decision function used to amalgamate individual member judgments into a group position influence the nature of the order effect and can vitiate it. Order effects, however, have a much weaker influence on conservatism than does message informativeness [12].

3. *Sequential versus simultaneous presentation.* Often in a MIDS, a choice exists as to whether to transmit information all at once or item by item (e.g. the use of a single aggregated report versus a set of individual reports). Our results indicate that significant differences in processing behavior occur when messages are received and processed simultaneously as opposed to sequentially with both individuals and groups. In some cases sequential information is processed more conservatively than simultaneous information and vice versa. The nature of these differences seems to depend in part on the perceived conditionality of the information sources, the specificity of the task, and the diagnostic element of the messages. Psychological experiments have similarly found that sequentially received and processed messages result in different judgments from simultaneously received and processed messages, partly due to an 'inertia effect' [20, pp. 704–706].

4. *Favorability.* Messages are generally either favorable or unfavorable. If favorable messages are processed differently from unfavorable ones, then messages might be worded such that unfavorable messages sound more favorable, or vice versa, depending on which wording results in more nearly optimal processing behavior. The bank lending experiments found that unfavorable messages (e.g. bad credit rating) tended to be weighted somewhat more than favorable ones (e.g. good credit rating) by individuals and artifactual groups [12]. Interacting consensus groups, however, seemed to treat unfavorable messages as being significantly more informative than favorable messages.

Several behavioral and marketing studies suggest contradictions and extensions of the above findings. Hughes [7], in a series of marketing experiments using executives surmised that the nature of the behavior is dependent upon message reinforcement (i.e. when a hypothesis about an object is favorable, a favorable message induces a greater positive change in attitude than an unfavorable message will induce a negative change, and vice versa).

5. *Objective versus subjective sources.* Information sources can be classified as being either objective or subjective, depending on the nature of the *data* and its source. Objective sources, for example, would be conditional probabilities known by the decision maker to be based on relative frequency data obtained from historical records and outputted from a computer. Subjective sources are known by the decision maker to have no historical data base; the conditional probabilities are subjectively determined and provided by another decision maker.

Messages received from subjective sources seem to be processed more conservatively than those from objective sources. This difference in behavior might be related to the difference in perceived credibility between these two types of sources. While these results have some psychological support, Hughes [7] found that the subjectivity of the messages was not considered by businessmen when acting on messages.

6. *Source complexity.* The complexity of an information source can be

defined by: (a) the number of possible messages emanating from it, and (b) the degree of asymmetry of the conditional probabilities relating the messages to the hypotheses. Examples of symmetry, for a given diagnostic element, are provided in Table 2. Symmetry, in a binary information source, for example, can be defined as the difference between the likelihood ratios of the two data states, and can vary from 0 (perfectly symmetric) to ∞ (maximally asymmetric).

TABLE 2.

	D_X	$D_{X'}$
H	90%	10%
H′	10%	90%

$L_X = 9$
$L_{X'} = 9$
$S = 0$

	D_X	$D_{X'}$
H	63%	37%
H′	7%	93%

$L_X = 9$
$L_{X'} = 2.5$
$S = 6.5$

	D_X	$D_{X'}$
H	45%	55%
H′	5%	95%

$L_X = 9$
$L_{X'} = 1.7$
$S = 7.3$

Information sources in real MIDSs are characteristically 'complex'; that is, there are many possible messages, and the likelihood functions relating the messages to the hypotheses are generally asymmetric. How complex or to what degree of accuracy and detail ('fineness') should an information source be specified to? Stated another way, should the data be made symmetrical and condensed or 'coarsened'? Normatively speaking, the more accurate and detailed the information the better, but realistically, 'symmetrizing' and 'coarsening' could result in improved assessments and information processing behavior.

In the bank lending experiments, messages from more complex information sources were processed more conservatively by individuals and there was less variance among the responses than with messages from simpler information sources (groups were not tested). Two explanations have been offered for this phenomena: (1) Symmetric sources were perceived as being more credible or trustworthy than asymmetric sources; (2) the behavior evinced was an artifact of the heuristics which decision makers employed in processing the information. That is, most individuals in attempting to intuitively weight the informativeness of an information source in revising their judgments do not consider the likelihood ratio associated with a given message in a manner similar to Bayes' Theorem. Instead they appear to focus on a single conditional probability, the one relating the message with the hypotheses in question.

7. *Format of presentation.* The format used to specify the informativeness of an information source can affect information processing behavior. To illustrate, consider the different ways in which the data from the information sources in the bank lending experiment could be presented (Tables 1 and 3). Only two conditional probabilities are needed, one associating any given message to each hypothesis. Thus, there exist four such possible presentations:

(1) $P(D|H), P(D|H')$ (likelihood ratio of message);

(2) $P(D'|H), P(D'|H')$ (likelihood ratio of alternate message);

(3) $P(D|H), P(D'|H')$ (conditional probabilities of 'correct' messages associated with each hypothesis); and

(4) $P(D'|H), P(D|H')$ (type I and II errors respectively). Theoretically there should be no difference in the way messages from each of these presentations are processed; but this is not necessarily so, psychologically. Consider, for example, the processing of message D. By using representation (1), the predictor (manager) receives the likelihood ratio directly, presumably facilitating intuitive processing of information more nearly in accordance with Bayes' Theorem. The other presentation formats provide different 'anchor points' which, if used directly, would result in suboptimal probability revisions, or require additional manipulations to obtain the likelihood ratio.

TABLE 3.

	D_X	$D_{X'}$
H	X	
H′	X	

	D_X	$D_{X'}$
H		X
H′		X

	D_X	$D_{X'}$
H		X
H′	X	

	D_X	$D_{X'}$
H	X	
H′		X

In the bank lending situation the various formats have been studied using both simple and complex information sources. Only slight to moderate differences in behavior occurred among the different formats. Training in the use of likelihood ratios for revising beliefs, as in Bayes' Theorem might, however, facilitate optimal behavior, especially in cases where binary hypotheses are involved.

Individual attributes and attitudes

1. *Perceived source credibility.* Although an individual may have verbalized his degree of belief in a hypothesis by means of a subjective probability, it makes sense to talk about his degree of belief or confidence concerning the accuracy of its numerical value.

Research in communication and persuasion indicates that one possible factor influencing the confidence associated with a probability judgment is an individual's perception of the credibility of the message he receives. Conservatism may derive then, in part, from an individual's disbelief in the datum itself or the source from which it came.

The results from the bank lending experiments showed that subject's perceived truthworthiness of the messages and sources (which were objective) did not influence the observed tendencies towards conservatism in either individuals or groups [10]. Other researchers [16], however, have shown a direct relationship

between an individual's revised probabilities and the extent to which he questions the trustworthiness of the information source. In fact, it has been observed that when source credibility is very low, people tend to become overly radical rather than conservative in their probability revisions [17]. The lack of significance in the bank lending experiments may have been due to the measurement instrument. Further research is suggested.

2. *Risk attitude*. One partial explanation of conservatism may stem from an individual's or a group's risk taking proclivity. That is, conservatism may be a function of risk-taking, and consequently individuals with high risk-taking propensities tend to be less conservative than those with low risk-taking tendencies. Using the standard Kogan and Wallach 'Choice Dilemma Instrument' (CDI) [8] as a measure of risk attitude, it was found that information processing behavior was not affected by the risk attitudes of individuals or of groups. These results should be interpreted cautiously on two counts: First, the CDI is an increasingly questionable measure of risk taking propensity [9]. Second, the majority of evidence in the psychological literature argues against the existence of risk taking propensity as a generalized characteristic of people. A person's previous learning experiences in specific risk-taking settings seem much more important than his general personality characteristics [19].

3. *Problem familiarity*. In making joint decisions, executives generally are or have an opportunity to become familiar with the problem beforehand, but in some cases do not. Without prior familiarization or experience a decision maker might not want to commit himself to a judgment until he had ample time to ponder it. This is especially so if the problem were either complex or important. With a complicated problem, the decision maker might feel he wanted an opportunity to weigh up all the evidence. With an important problem, even a simple one, he might be reluctant to make a hasty decision. In both cases he might feel inclined to give an initially cautious response, meaning "Don't know" or "Not yet ready to decide". After he had time to study the problem, however, his initial caution may vanish. In psychological studies of risk attitude further familiarization with the risk-taking problem led to an increase in riskiness [4]. Such an effect might conceivably occur when processing information and thus could counterveil the conservatism phenomenon. Our findings show, however, that individuals and groups with prior problem familiarization do not process information significantly differently from unfamiliarized individuals and groups. Strub [22], however, found that highly experienced decision makers in the military sphere, were clearly more nearly optimal than inexperienced decision makers. But Slovic's observations of investment analysts indicate that experienced and familiarized decision makers may be less accurate than those with less experience and familiarization [18].

Social interaction

The decision maker in an organization processes information in a wide variety

of social situations including virtual isolation, conference telephone calls, committee meetings, and numerous others. The characteristics of the information processed (e.g. its diagnosticity, complexity, etc.) are similarly diverse. The question arises as to whether the decision unit (viz, an individual or group) processes the various forms of information similarly in the numerous social situations encountered. If the social situations and the nature of the information evoke a change in processing behavior (ultimately reflected in the final judgment), it may then be desirable to adapt the social situation to the information which is to be processed, or, conversely, adapt the form in which (or the media through which) the information is presented to the nature of the decision unit. It is thus important to examine systematically the effects of social interaction on information processing behavior in individuals and groups. Such a study was performed using the bank lending situation [13]. In this study social interaction was defined in terms of social exchange and social environment. Social exchange refers to the degree of communication exchange or feedback from person to person, while social environment is concerned with the isolation or presence of others. The results found were: (1) Individuals, artifactual groups, and interacting consensus groups were not affected by social interaction in the processing of messages of medium and high informativeness. (2) However, with messages of low informativeness both individual and group responses were affected by social interaction. At this level of informativeness individuals working alone without discussion and groups who functioned in visual isolation were the best information processors.

The implications inherent in these findings are in certain respects contrary to the typical assumptions which underlie certain management practices. Committees are usually organized for judgmental and decision making purposes for those situations in which uncertainty is high and the consequences are great. The rationale, for such action, at least in part, is based on the "two heads are better than one" adage. In government, for example, judgments and decisions involving matters of national security are usually entrusted only to committees, as the conditions under which such judgments and decisions are made are generally highly charged and fluid, and can have catastrophic consequences. These findings question the assumption upon which this practice is based, as individuals acting alone, made better and less extreme judgments with highly uncertain information. The results also question the value of visual communication devices for probabilistic judgments (e.g. such as the newly developed picture-phone) in lieu of verbal or written communication which require no visual contact. Operationally, then, the very same message through different media (such as written media, a telephone, T.V.) may indeed have a significantly different impact on the audience (in our case, decision maker), partially as a result of social interaction.

It would be well to mention that there is no evidence in the psychological literature regarding the effects of social factors on the type of human judgmental behavior that we are concerned with. A fuller understanding of the processes

involved in the interaction of social factors and information characteristics warrants further study.

Group decision function

When a group judgment is required, some rule or "group decision function" must be used to amalgamate individual members' judgments into a consolidated group position. There exists a number of such rules, which can be conveniently categorized into four types: (1) informal, (2) Delphi, (3) mathematical, and (4) voting rules. Informal techniques (letting the group arrive at a consensus via discussion) while admittedly *ad hoc*, are nevertheless the most common in practice. The Delphi technique is probably the most widely known formal method for eliciting and refining the judgments of a group of people. Delphi does not permit any direct contact among the people involved. It is usually done operationally through an interactive questionnaire in which each person, on a given cycle, purveys his judgment based on a summary of the judgments of each person from the previous cycle.

There exists a variety of mathematical aggregation techniques [2]. However, only a few such rules were examined in the bank lending experiments. One *ad hoc* rule is to average the individual group member's likelihood ratios implied by his judgments. Another mathematical rule is a Bayesian amalgamation procedure proposed by Roberts [14]. In brief, the group posterior distribution is determined by taking a weighted average of each individual member's posterior distribution. The weights are formally revised after each message according to each individual's assessment. Numerous voting rules also exist as a means of aggregation. Only majority rule was examined. For three member groups this is the median probability response value.

Results show that the rules result in different judgments (no one rule being clearly superior to any other) and that only the judgments resulting from a few of the rules tested were affected by order of message presentation, favorability of the message, etc. Much work needs to be done in this area. Paralleling our efforts here is a large body of work now being done on this topic by social psychologists and behavioral decision theorists.

RELATIONSHIP OF FINDINGS TO REAL WORLD

The results of these and other experiments, while not conclusive, indeed, show that man has difficulty both individually and collectively in processing probabilistic information. Many factors appear to affect his performance in various ways, some more predictable than others. Moreover, much of man's biases and fallibilities found in the bank lending studies replicate and extend those found in psychological studies. To generalize beyond this, however, is risky. Our experiments, although more realistic than the psychological experiments, were not reality. Some obvious potential differences that often exist in real-world situa-

tions and not found in the bank lending studies are: (1) conditional dependence among information sources, (2) nonstationarity of the environment (e.g. the hypotheses change, the messages change, or the conditional probabilities relating the messages to the hypotheses change, etc.), (3) relative undiagnosticity of the real information sources, and (4) unreliability or cascaded inference structures (inferences involving a hierarchy of several levels) [27].

Nevertheless, there exists an abundance of evidence in the real-world indicating that the above biases and difficulties in processing information observed in the bank lending studies and the psychological laboratory persist outside the laboratory in business, military, and societal decision making situations [12]. Consider order effects: note its relationship to the adage, "First impressions are most lasting". How many times have we been in the miserable position of attempting to overcome a first bad impression (primacy effect); or, being unfairly judged for a merit increase or promotion based on events and performance over the past few weeks rather than over the entire review period (recency effect). Consider conservatism: this is the fundamental principle of the accounting profession! In marketing, salesmen's sales forecasts usually, if not virtually always, underestimate actual sales; in R&D, project completion time is almost always severely underestimated.

Vol. 4
1976

OPTIONS OPEN TO MIDS DESIGNER

Humans need help in making sound judgments. While training will mitigate errors and biases to some extent, it will not eliminate them entirely, either because of the complexity of the many different situations encountered, fatigue, or just the sheer infeasibility or impossibility of developing a training program that will achieve errorless results. Managers, information specialists, and operations researchers must therefore identify and account for human biases and deficiencies in order to design better MIDSs. What are some of the options open to them?

1. *Incorporate in the system counter-biases or eliminate conditions which create human biases and inconsistencies.* For example, with respect to order effects: (1) Control the timeliness with which data is presented to the information processor. For instance, the information system specialist can analyze the delay time patterns (i.e. elapsed time from the original data to its presentation to the decision maker) to determine any bias that would cause one kind of data to be systematically presented before another and to analyze forms and reporting structures for any obvious sequence biases. Here attempts could be made to synchronize data delay patterns and to reorganize forms and reports. (2) Alert information processors to the possible impact of order effects on behavior and suggest they re-evaluate all relevant data in that light prior to making a judgment (of course, here, one runs the risk of imparting new biases).

2. *Exploit human biases and irrationalities.* Consider, for example, the societal decision problem of vehicle accident prevention. The actual (objective) probability of detection of a speeding violation is exceedingly small; in fact, it is a rare event. Yet people's beliefs regarding the probability of detection might be surreptitiously inflated by withholding information (keeping such data confidential), providing misinformation (propaganda), repeated or redundant information, or by simply dramatizing the consequences (resulting in a violation of Savage's independence of beliefs and tastes axioms of decision theory [5] which humans tend to violate). Similar strategies could be advantageously employed for problems in pollution detection, nuclear reactor safety, etc.

3. *Decision aids.* Automate the information aggregation process (let a mathematical model and machine do it). Bayesian Probabilistic Information Processing (PIP) systems and regression models already have been proposed and successfully used in a variety of real decision making contexts [20]. Such models eliminate an individual's inconsistency, hence seem appropriate for stationary environments and perhaps for nonstationary environments as well.

Another approach is to incorporate human fallibilities directly into the decision model. For example, inspector errors have been incorporated directly into mathematical models for determining optimal sampling plans.

4. *Incorporate flexible and adaptable control systems.* Most judgments and decisions will continue to be made subjectively by humans. To assure better judgments and more efficient use of the organization's resources, imbed in the planning and control function (a) a 'proper' reward function to assure honest judgments, and (b) a feedback system to foster rationality and optimality through learning [26].

CONCLUSIONS

Considerably more research must be done on human information processing behavior particularly in realistic environments with emphasis on understanding the heuristics people use in arriving at a judgment both individually and collectively. It is only recently that any effort at all has been made in trying to determine what rules people use in forming and revising their beliefs [23]. No longer can the psychologist remain in his laboratory and no longer can the information specialist and operations researcher ignore human behavior if MIDSs are to be more than marginally successful.

REFERENCES

1. ACKOFF RA (1967) Management misinformation systems. *Mgmt Sci.* **14** (4), 147–156.
2. DEGROOT MH (1974) Reaching a consensus. *J. Am. Statis. Ass.* **69**, 118–121.
3. EBERT RJ (1970) Methodology for improving subjective R&D estimates. *IEEE Trans. Engng Mgmt* **17**, 108–116.

4. FLANDERS JP and THISTLETHWAITE DL (1967) Effects of familiarization and group discussion upon risk taking. *J. Personality & Soc. Psychol.* **5,** 91–97.
5. FORRESTER J (1961) *Industrial Dynamics.* MIT Press, Cambridge.
6. GOODMAN BC (1972) Action selection and likelihood ratio estimates by individuals and groups, *Organ. Behav. & Hum. Perform.*, **7,** 121–141.
7. HUGHES GD (1971) *Cognitive models and individual determinants of changes in subjective probability.* Delivered at the International Meeting of the Institute of Management Science, Washington, D.C.
8. KOGAN N and WALLACH MA (1964) *Risk Taking: A Study in Cognition and Personality,* Holt, Rinehart and Winston, New York.
9. MACKENZIE KD (1971) An analysis of risky shift experiments. *Organ. Behav. & Hum. Perform.* **6,** 283–303.
10. MASON RO and MOSKOWITZ H (1972) Conservatism in information processing—implications for management information systems. *Decision Sci.*, **3** (4), 35–55.
11. MJOSUND A (1972) The synergy of operations research and computers. *Ops. Res.* **20** (5), 1057–1064.
12. MOSKOWITZ H & MASON RO (1973) Accounting for the man/information interface in management information systems. *Omega,* **1** (6), 679–694.
13. MOSKOWITZ H and SILVA WT (1973) *Effects of social interaction on human probabilistic inference.* Krannert Institute Paper No. 397, Purdue University.
14. ROBERTS HV (1965) Probabilistic prediction. *J. Am. statis. Ass.* **60,** 50–62.
15. SAVAGE LJ (1954) *The Foundations of Statistics.* Wiley, New York.
16. SCHAEFER RE and BORCHERDING K (1973) A note on the consistency of two approaches to incorporate data from unreliable sources in bayesian analysis. *Organ. Behav. & Hum. Perform.* **9** (3), 504–507.
17. SCHUM DA, DU CHARME WM and DEPITTS KE (1973) Research on human multistage probabilistic inference processes. *Organ. Behav. & Hum. Perform.* **10** (3), 318–348.
18. SLOVIC P (1972) Psychological study of human judgment: implications for investment decision making. *J. Fin.* **27,** 779–799.
19. SLOVIC P, FLEISSNER D and BAUMAN WS (1972) Analyzing the use of information in investment decision making: a methodological proposal. *J. Bus.*, **45,** 283–301.
20. SLOVIC P and LICHENSTEIN S (1971) Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organ. Behav. & Hum. Perform.*, **6,** 649–744.
21. SOUDER WE (1969) The validity of subjective probability of success forecasts by R&D project managers. *IEEE Trans. Engng Mgmt,* **16,** 35–49.
22. STRUB MH (1969) Experience and prior probability in a complex decision task. *J. Appl. Psychol.*, **53,** 112–117.
23. TVERSKY A and KAHNEMAN D (1973) *Judgment under uncertainty: heuristics and biases.* Proceedings of the Fourth Research Conference on Subjective Probability, Utility, and Decision Making. Rome.
24. VON NEUMANN J and MORGENSTERN O (1947) *Theory of Games and Economic Behavior.* 2nd ed. Princeton University Press, Princeton.
25. WAGNER HM (1971) The ABC's of OR. *Ops Res.* **19** (6), 1271–1281.
26. WINKLER RL (1969) Scoring rules and the evaluation of probability assessors. *J. Am. Statis. Ass.* **64,** 1073–1078.
27. WINKLER RL and MURPHY AH (1973) Experiments in the Laboratory and the Real World. *Organ. Behav. and Hum. Perform.*, **10,** 252–270.
28. WOODS DH (1966) Improving estimates that involve uncertainty. *Harv. Business Rev.* **44,** 91–98.

ADDRESS FOR CORRESPONDENCE: *Professor H. Moskowitz, Krannert Graduate School of Industrial Administration, Purdue University, Lafayette, Indiana 47907, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Some Properties of Pareto-optimal Choices in Decision Problems

HC CALPINE

A GOLDING[1]

(*Received February 1975; in revised form August 1975*)

Estimates are derived for the expected number of Pareto-optimal choices when n candidates are compared with respect to q qualities. The number of comparisons needed to find the Pareto-optimal choices, and the possible advantages of sub-optimization procedures, are briefly discussed.

ol. 4
976

MANY PRACTICAL one-time decision problems, for which the primary data-gathering and evaluation phase is complete, can conveniently be formulated to management in a 'decision matrix'. In this the row labels are alternative 'candidate' choices open to the decision-maker (e.g. on allocation of funds, choice of equipments, specification of military forces), and the column labels are particular 'qualities' considered to be relevant to the decision (e.g. performance in different situations, costs in particular resources, returns per unit expenditure etc). The entries in the matrix represent the quantitative evaluation of each 'candidate' (row) in respect of each 'quality' (column). It is usually desirable that the column labels represent 'independent' qualities in the colloquial sense of the term, i.e. the entries in one column are not predictably related to those in another by the inherent structure or the formulation of the problem. Except where otherwise stated it is supposed in what follows that preference is for candidates having the higher entries in the columns, i.e. the general aim is maximization.

A candidate is 'dominated' if there is another candidate which excels it in one or more qualities and equals it in the remainder. By the usual elementary reasoning the decision-maker can discard from consideration all dominated candidates. It may be that there is a 'dominant' candidate (or possibly more than one with identical entries) such that all other candidates are dominated by

[1] Formerly of the Directorate of Naval Operational Studies, Ministry of Defence, London SW1.

it/them. If there are such, they are clearly preferred choices. More commonly the decision-maker is left with a group of undominated candidates, some of which are preferred in some qualities and some in others. This is the so-called "Pareto-optimal" (PO) set.[2] There are many methods discussed in the literature (e.g. weighting of columns, maximizing expected utility, minimax regret etc.) of narrowing the choice further among the (sometimes numerous) members of the PO set, but it has been the authors' experience that while the arguments leading to the PO set are generally accepted as non-controversial, the ideas implicit in the more selective methods of choice are not readily agreed among decision makers. Interest in the PO set may be either as a sufficient indication of near-optimal choices (the final choice then being made 'on other grounds' e.g. convenience), or as a reduction in the number of candidates to which it is necessary to apply more elaborate analytical techniques of choice. It seemed worthwhile to explore the process of finding the PO set in rather more detail to see how good and convenient a sieve it is. For example it is intuitively clear that if the number of independent 'qualities' is increased, the PO set eventually includes all the candidates, and there is then no sieve effect at all. Can numbers be put to the selectiveness of the sieve as a function of the number of candidates and qualities? Then again, it often occurs that the number of 'candidates' generated by the combination of various possible features is very large. It is therefore of interest to know whether "Pareto optimization" can be broken down into more easily handled "Pareto sub-optimizations" over convenient groupings of candidates. These topics will now be taken up in turn, starting with the expected number of candidates in the PO set of a decision matrix.

Vol. 4
1976

THE AVERAGE NUMBER OF PO CANDIDATES IN A MATRIX OF n CANDIDATES AND q QUALITIES WITH RANDOM ENTRIES

Consider first the apparently very special case in which all the entries in the matrix are random numbers uniformly distributed over the range 0 to M. Attention is first focussed on the final (q^{th}) column. Arrange the rows so that the entries in the q^{th} column are in decreasing order of magnitude. By the randomness of the entries, the probability of an arbitrarily selected row being r^{th} in the order is $1/n$. Let $p(n,q)$ be the probability that a row, arbitrarily chosen from n rows, is PO with respect to q qualities. Consider the r^{th} row. The ordering ensures that this row is not dominated by any row below it and also that it exceeds no row above it in the q^{th} quality. Hence a necessary and sufficient condition for the r^{th} row to be PO is that it is PO among the first r candidates with respect to the first $(q-1)$ qualities. Thus the probability of a row being r^{th} and PO is $p(r,q-1)/n$. The probability of an arbitrarily selected row being PO is

[2] For formal completeness dominant candidates if any will subsequently be considered as included in the description "PO set".

$$p(n,q) = \sum_{r=1}^{n} p(r,q-1)/n$$
$$= \{p(n,q-1) + (n-1)\,p(n-1,q)\}/n.$$

An expression for the average number a(n,q) of candidates in the PO set follows by substituting $p(n,q) = a(n,q)/n$

$$a(n,q) = a(n,q-1)/n + a(n-1,q)$$

As $a(n,1) = a(1,q) = 1$, $a(n,q)$ can be calculated recursively. Some typical results are shown in Fig. 1. For the discussion of general properties it is convenient to express $a(n,q)$ in closed form. A good approximation is given by the following expression:

$$a(n,q) \sim 1 + \log n + (\log n)^2/2! + \ldots\ldots$$
$$+ (\log n)^{q-3}/(q-3)! + \gamma\,(\log n)^{q-2}/(q-2)! + (\log n)^{q-1}/(q-1)!$$

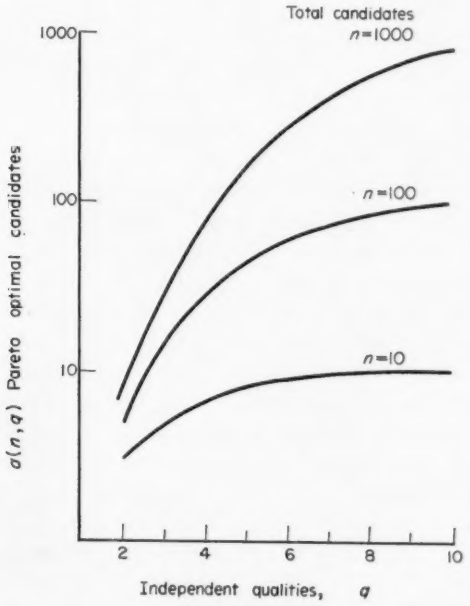


FIG. 1. *Expected number of PO candidates from total* n *assessed for* q *independent qualities.*

where γ is Euler's constant ($\sim 0{\cdot}5772$). An analytical derivation has been obtained and is available on request. On the scale of Fig. 1, estimates by this formula are indistinguishable from the recursive calculations.

GENERALIZATION OF THE RESULTS

A convenient way of thinking about characteristics of Pareto optimization and making approximate calculations is to use a geometrical representation. The entries in the matrix can be regarded as the specification of q orthogonal co-ordinates for each of n (candidate) points. For ease of exposition suppose all entries to be positive. Consider the q planes parallel to the co-ordinate axes through a particular candidate point. These divide the co-ordinate space into 2^q separate regions, one of which contains the origin. By analogy with the easily visualized two and three dimensional cases let us call the region so distinguished the 'box' associated with the particular candidate point. Then the definition of Pareto-optimality for a candidate is equivalent to the statement that its representative point is not included in the 'box' associated with any other candidate. Using the original assumption of a uniform distribution of matrix entries, the probability of this occurring can be estimated, as can other features not considered in this paper. However, the geometrical representation suggests that for $a(n,q)$ a powerful generalization is possible. The relation of 'inclusion' is not changed by any order-preserving transformations along the co-ordinate axes. The result obtained for uniform random entries will therefore apply to any distributions of column entries which can be reached from a uniform random distribution by order-preserving transformations. Such a transformation can be found for any continuous probability density function. The result therefore applies in many practical cases. Provisos are firstly that the statistical correlation between different columns is low, i.e. the 'colloquial' independence of qualities initially postulated is reflected in a statistical near-independence. Positive correlation between the quality ratings tends to decrease the average number of PO points, and negative correlation to increase it. The second proviso is that a complete order of merit can be established in each quality. The effect of 'ties' is usually to decrease the number of PO points, but unless the number of 'ties' is large, as can occur with grouped data, the change is small. The generalization that the value of $a(n,q)$ applies to all practical data which meet these conditions can also be deduced from the derivation of the recurrence relation in the previous paragraph. It will be noted that this requires only independence between columns of the ordering of column entries (so that $p(r,q-1)$ is independent of the ordering in the q^{th} column) and makes no reference to magnitudes. It will therefore apply to distributions of magnitudes which are otherwise arbitrary.

The following qualitative conclusions are suggested by Fig. (1). The sieve formed by the Pareto-Optimization process is reasonably selective only for small numbers of 'qualities'. The *relative* reduction in number of candidates increases with the total number of candidates.

e.g. for three independent qualities the ratios of numbers in the PO set to total number of candidates is approximately

Vol. 4
1976

0·506 for 10 candidates
0·143 for 100 candidates
0·029 for 1000 candidates

So the sieve is certainly useful for the not uncommon case where a relatively small number of 'qualities' is concerned, and where combinatorial processes generate very large numbers of 'candidates' for consideration.

For practical purposes the expected number of points in the PO set when n candidates are compared with respect to q independent qualities is well approximated by the expression given at the end of the preceding sections.

PARETO SUB-OPTIMIZATION

Taking up the question of how to handle large-scale problems, consider as an example the case where a diverse set of 'capabilities' have to be simultaneously provided, each at a specified level. For each capability there is an available choice of several different equipments, but for simplicity suppose that the capabilities are so defined that an equipment is only applicable to a single capability. In practice this is usually possible, for example a separate 'common services' capability may be defined for multi-use equipment. A candidate consists of a choice of one or more equipments to provide every capability and so if the number of equipments is at all large, the combinatorial features generate very large numbers of candidates. However the total number of comparisons between rows needed to identify all the PO candidates is normally much less than the total possible number of comparisons between rows. Thus the following procedure could be used: Compare the first two candidates and if either is dominated discard it. Next compare the undiscarded candidates with the third candidate and discard any dominated candidate. Then introduce the fourth candidate and so on. After $n-1$ stages the PO set is determined and the undiscarded candidates after the r^{th} stage are the PO set for the first $r+1$ candidates. It follows that the average number of comparisons at the r^{th} stage is not more than $a(r,q)$ so the average number of comparisons to find the PO set for n candidates is less than

$$T(n,q) = \sum_{r=1}^{n-1} a(r,q) = na(n-1,q) - T(n,q-1).$$

An example of this upper limit to the average number of comparisons required for the previously quoted examples for three qualities is

0·632 of all possible comparisons for 10 candidates
0·203 „ „ „ „ „ 100 „
0·045 „ „ „ „ „ 1000 „

The upper limit to the average proportion of all comparisons needed is equal, for large q, to the proportion of candidates which are PO and therefore

approaches unity. For $q = 1$ the proportion of all comparisons needed is twice the proportion of candidates which are PO.

Some further reduction may be possible if the candidates are partitioned in some convenient way into separate groups. Any candidate which is PO among all the candidates must be PO (or dominant) among any selection of candidates. Therefore a possible procedure is to find first the Pareto sub-optimal (PSO) set for each group and then 'represent' each group by its PSO set in a final comparison. However in this simple form the partitioning procedure does not usually yield any great reduction in computation. This is partly due to the fact that the PSO sieve is less selective for the smaller number of candidates in a group, and partly to the final selection having to be made from a class of candidates which has a relatively large proportion of PO members, and in which therefore a relatively larger number of inter-row comparisons have to be carried out.

In the same type of problem, a more economical decomposition can be achieved if it is possible to sub-optimize for the provision of each capability separately before combining the choices. An interesting case occurs when the entries for equipments are additive in each 'quality column'. For example the 'qualities' might be costs in various different scarce resources, such as capital cost, R & D cost, manpower, etc. The interest would then be to find those candidates which provided all the specified capabilities at costs which in a general sense were 'as *low* as possible' in all the categories. The example is therefore a minimization rather than a maximization problem, but the same general arguments apply. The PO set now comprises the candidates whose costs are not greater than those of any single competitor in respect of every category of cost and are less in at least one category of cost. Suppose then that the problem is broken down by capabilities, and the choice of equipments to provide a single capability is considered as a sub-problem. A sub-matrix can be written down which records the various 'costs' of these equipments. Suppose further that by any process whatsoever one equipment (say e_c) is selected from the sub-matrix to 'represent' that particular capability in the specification of a candidate for the main matrix (in which each candidate has to provide all capabilities simultaneously). Then it is a necessary (but not of course sufficient) condition for the candidate using this 'representative' equipment choice to be PO in the main matrix, that the 'representative' equipment (e_c) should be PSO in the sub-matrix. (For if it is not, there will be at least one other equipment (say e_c') in the sub-matrix whose costs are no greater in any category and are less in at least one category. If therefore e_c' is substituted for e_c in specifying the candidate in the main matrix, the costs for that candidate will not be increased in any category and will be reduced in at least one category. So the candidate using e_c could not have been PO in the main matrix.) Therefore in the specification of candidates for the full problem *only those equipments which are PSO in the corresponding capability sub-matrix need be considered.*

Vol. 4
1976

It will be noted that while the above has been proved for the case where entries in a column are the sum of the entries for the various equipments used by a candidate, the ordinal nature of the Pareto-optimality concept means that the same argument applies whenever the column entries in the full matrix are built up from the equipment entries in the sub-matrices by a combination rule such that the entry in the full matrix increases monotonically with each of the representative equipment entries. For example if the 'qualities' were probabilities, the combination rule might be multiplicative, but it would still be true that only equipments which are PSO in their capability sub-matrix need be considered in specifying candidates for the full matrix.

As a numerical example, consider the case where there are 10 different possibilities for providing each of three capabilities, the candidates being judged on the basis of three qualities. The total number of rows in the complete decision matrix is 1000 and the expected number of PO candidates is about 29. Direct identification from the full matrix will involve about 13,000 row comparisons. Sub-optimization for each capability will provide about 5 possibilities for each capability which will lead to a final matrix of about 125 rows. The total number of row comparisons involved is difficult to estimate but appears to be of the order of 1300.

It is concluded therefore that it may sometimes be useful to simplify and reduce a large decision matrix by Pareto sub-optimization on the input data. This is particularly helpful when different people or organizations are concerned with providing the input data (e.g. on the various capabilities) in that it provides an easily understood 'selection rule' which can be used by all parties.

This work was part of a study undertaken in the Navy Department, Ministry of Defence, by whom Crown copyright is reserved.

ADDRESS FOR CORRESPONDENCE: *HC Calpine, 1 Fairfax Road, Chiswick, London W4 1EN, UK.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

The Communication of Control in an International Company: Methods of Investigation[1]

MICHAEL G DORRELL

MICHAEL Z BROOKE

(*Received March 1975; in revised form October* 1975)

This article describes the methodology developed for the objective analysis of communication, in terms of its control content, in one international company. The research project examined detailed aspects of control and communication practices in the firm with some comparative material from other companies. This article is intended as a contribution both to the understanding of a complex communication system, and to the measurement of control in an international group of companies. For reasons of space an account of the mechanics of the analysis has been omitted.

THE CHANNELS OF COMMUNICATION

EVERYONE communicates. Lip service is paid to its importance, but few positive steps are taken for its improvement. When communication is very faulty it is instantly recognisable as such; when it works well it is ignored. In effect, unless communication is extremely bad, businessmen tend to treat the matter as unimportant. This is especially so of international business, where the complexities are the more intricate for its international aspects. These tend to diminish even more the pressure to improve communication. The pressures only become irresistible when the communications are at the point of breaking down.

None of this helps the researcher. He quickly discovers that an interviewee's

[1] The methods here described were evolved by Michael Dorrell when a researcher with the International Business Unit in the University of Manchester Institute of Science and Technology. The Unit has fifteen researchers investigating various aspects of the management process in international firms. In addition to support from Business and the University, the Unit has had substantial Grants from the Social Science Research Council and the Scandinavian Institute for Administrative Research.

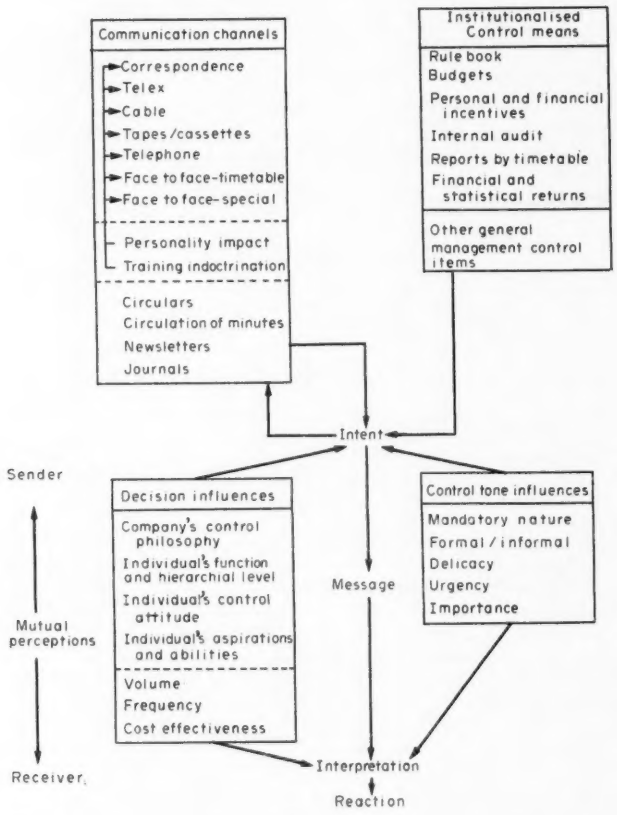


FIG. 1.

perceptions about communication reflect the prevailing attitudes; and these, at a superficial inspection, can produce some extremely illogical situations. One must be careful about predicting communication behaviour on the basis of generalisations drawn from perceptions.[1] Hence the need for more objective data. Figure 1 summarises the treatment of communication in this research. The thirteen channels which are theoretically available for use by the Company have been identified as those listed in Fig. 1.

Four of these need further explanation.

Face to face communication refers to visits. The distinction is made between visits whose origins lie in a timetable, and those which are undertaken for special reasons. The former would include board meetings and regular tours, the latter might be for specific problem solving purposes.

By *personality impact* we refer to the conscious move by a superior to take up a matter normally dealt with by his subordinate to make the communication

Vol. 4
1976

of that matter to a third party more effective. This item contains a strong controlling or influencing aspect, and as such is one of the several arbitrary placings for the researcher to decide when making artificial separations of a business system for purposes of investigation. This is one of two abstract channels of communication included in our list, which can both theoretically be applied to any one of the preceding seven channels. Its frequency of use as a technique for communicating control justifies its separation into an identifiable channel of communication.

Training indoctrination, the second of the abstract channels, refers more to the art than the mechanics of communicating. This is usually, though not exclusively, used in the context of communication between Head Office and an expatriate receiver overseas. Either the message content or the degree of influence or both are deliberately camouflaged by the sender. The receiver is able to interpret both content and influence correctly in the light of prior agreement or instruction. This intended deception of third parties poses obvious problems of interpretation for the researcher.

Theoretically all the *channels* listed in Fig. 1 can be used to transmit information concerning each of the institutionalised control *means*. In practice certain well trodden routes will evolve and some theoretically possible routes will never be used. In our principle of working from the most elementary combination of variables in their lowest common denominators we must consider all possible combinations within our framework for analysis.[2]

Each combination of channel and means which exists in practice has been selected by the sender to transmit a message in such a way that the receiver's interpretation and consequent reaction will be that intended by the sender. In this simple progression—control means, selection of communication channel, intent, message, interpretation and reaction—the sender is seen as manipulating the communication system in order to maximise the effectiveness of his message.

The research, on part of which this article is based, attempts to show the extent to which the implementation of a control system is dependent on the communication system which serves it, and to isolate and examine the influential sub-systems which complicate and confuse the simple progression described above.

For the purposes of Fig. 1, we have isolated two types of influences which will contribute with varying degrees of emphasis both to the sender's intent and consequent choice of channel and to the receiver's interpretation. The complexity of this process is compounded because the individual influences may

[2] Cards were used during the interview to present to the interviewee the full range of theoretical possibilities in his use of control means and communication channels. From these cards he was asked to select the range and combination of alternatives which he used; he was also asked to answer specific questions on the cards selected and their combinations. The interviewee showed relief when he realised that the sets of cards he was considering would be substantially reduced according to his experience with both channels and means.

The use of cards had an unexpectedly beneficial effect on the atmosphere of the interview.

have more or less impact on the sender than the receiver. Moreover, the perceptions which one person holds of the impact of these influences on the other—here called 'mutual perceptions'—may in certain instances have a marked effect on either the sender's intent or the receiver's interpretation.

The framework for treating the communication of control in this research, summarised in Fig. 1, thus presents a theoretical possibility of many thousands of variables and combinations of variables for the researcher to identify and examine. In practice, however, a sender will not always consciously select a particular strategy, rationally argued, for each occasion. Many situations will be repeats of earlier ones, whose treatment in terms of the matching of intent and interpretation has evolved often over many years to establish a tradition of stimulant and response. This applies not only for individuals who have grown with the situation, but also for the particular role incumbents who by 'training indoctrination' foster control and communication systems devised by others.

New situations may arise, or specific influences may become unusually dominant in a given situation. In such cases, the sender can and does make decisions and choices from the alternatives available to him based on the dominant influences of the situation in order to maximise the effectiveness of his message on the receiver.

The collection of data for communication analysis started very early in the project. The ideas for the objective measurement of communication developed from a suggestion from the head of the division during the first week of the research. He said "the best way of finding out what we do here is to read the letters". This seemed a daunting prospect, for there were many hundreds of files. In fact this undertaking was greatly simplified, and further developments were made possible, by an internal procedural practice where correspondence was concerned. In this division, every member was required to place an extra copy of all correspondence in a special file. These copies were numbered and periodically sorted into chronological order. All members were involved, even the technical representative of the division located in the laboratories. One could be sure that for a certain day, week or month every letter written within the division could be read. The exceptions to this were the very private and confidential letters of the head of the division, and some of the correspondence of the responsible director on the group Board. The extent to which these two omissions affected the results of the analysis will be assessed later.

Correspondence is but one channel of communication of the thirteen listed here, but it is undoubtedly the one which carries the greatest volume of messages and which is used most frequently. Moreover, with certain exceptions, the researcher could be certain that he had access to almost 100 per cent of the messages transmitted from Head Office to the foreign subsidiaries through this channel. Because the area of particular interest was that of control, the researcher could reasonably expect to find that message senders at Head Office would not

use correspondence exclusively when dealing with control items. Indeed of all aspects of business affairs control items, because of their very nature, would probably be as widely scattered throughout the available channels as any other aspect. This supposition proved to be correct and affected the research in two different ways.

The negative impact was that our 100 per cent cover of the measurement of control items through correspondence appeared to be less impressive when it was realised that of the total number of control items passing through the communication system by no means all would be found in correspondence. This fact emerged at the beginning of the research, but was more than outweighed by the positive impact which came to light as the research developed.

It soon became apparent that control items which were initially transmitted by other channels, such as telex or telephone, were confirmed later in correspondence. This was not the case for all messages conveyed by telex or telephone, but most control items were sufficiently important for this procedure to be followed. This fact had a profound effect on the research. The researcher could now apply part of the benefits derived by his 100 per cent cover of correspondence to the volume and nature of control items passing through other channels whose comprehensive measurability would otherwise have been much more in doubt.

The overriding difficulty in this research was always that the ability to measure objectively control items passing through all the channels varied from almost comprehensive cover (correspondence) to comparatively superficial cover (training indoctrination). The researcher can always use perceptions for the measurement of areas for which documentary evidence is not available or accessible, but his confidence in, and the comparability of, data gleaned from these two sources will vary considerably.

In this case study, perceptions have been used as measurements where no documentary evidence exists or can be studied and to complement those measurements of other areas which are based on documentary evidence. The limiting factor, therefore, on the strict comparability of measurements of the different elements of the communication system was the lack of documentary evidence for certain channels of communication. It was clear that this research could not produce detailed definitive statements with complete confidence on the workings of the case study communication system as a whole. This encouraged the adoption of the approach which can be described as piecing together the view by looking through different windows. In other words, the researcher used several approaches to his study of the communication system. The results of each of these would throw further light on both the individual element under study and the system as a whole. A prolonged period of experimentation with measurement techniques and methodology and the cross-checking of results gleaned from different measurements of the same data was undertaken. This led to the gradual development of techniques of measurement

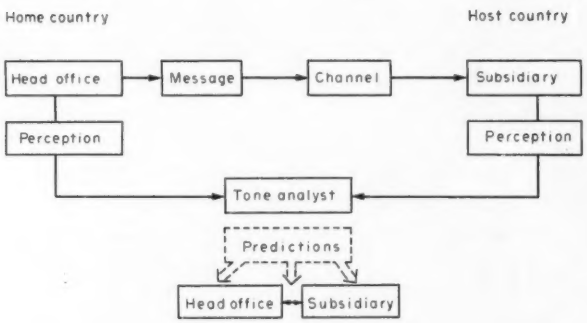


FIG. 2.

which made their varying contributions to the researcher's attempts to unravel the complexities of the communication system.

Experiments were made, for example, to measure objectively the content of face to face meetings between Head Office and Subsidiary personnel and the use made of the telephone. Personality impact and training indoctrination were similarly investigated. Descriptions of these experiments, however, are not included in this article, which concentrates on the development of correspondence analysis which eventually became the researcher's most valuable and reliable source of analysed data concerning the communication system.

Correspondence analysis was designed primarily to measure the control content of correspondence, which can be referred to as its *tone*.

Content analysis operates in a much wider context than does our 'tone' analysis. Much of the work of Budd, Thorp and Donohew [2], for example, is focused on the content of newspapers and magazines—and, in general, it is the mass communication media to which this approach is usually geared. Figure 2 depicts the process by which we have reduced the scope of this analytical tool to focus specifically on the tone in correspondence from Head Office to overseas subsidiary.

The correspondence was examined over a four year period and classified under the following headings:

1. Mandatory orders:
   Very strong suggestions from Head Office—outright orders—non-routine urgent requests. Outright orders were rare; nearly all were disguised to suit the situation and subsidiary in question. The analysis was designed to penetrate the disguise and record the intent.
2. Suggestions:
   Advice not asked for by the subsidiary company.
3. Advice:
   Information and advice asked for by the subsidiary—specifically in response to a query or request and follow-up correspondence to the original response.

4. Queries:
   Concerning errors or ambiguities in communication to Head Office.
5. Apologies:
   From Head Office for lateness in action promised or not answering letters, or for confusion over goods ordered or despatched.
6. Information:
   Not asked for by subsidiary and sent as a matter of routine.
7. Requests:
   From Head Office, of a relatively unimportant nature, together with apologies and complaints.
8. Unrequested suggestions and information.

SOME RESULTS OF THE ANALYSIS

The method described is suitable for analysing in depth the correspondence and, with certain qualifications, the communications patterns of an international group. Space does not permit a detailed discussion of the problems involved, but the method can be used with confidence provided the possible distortions are taken into account.

The scale of the exercise is such that we can use the results as a reference for considering a number of hypotheses about organisations, communications, centralisation, control and other such issues. The value of this reference work lies in its comparative objectivity. It is also possible to interpret more subjective perceptions, if not entirely accurately, at least in the context of what actually did happen.

For present purposes, we can examine the scope of the results achieved by the analysis by describing briefly six aspects ranging from broad general trends which can be gradually broken down in order to investigate the origins of these trends. It is in this process of progressing from the general to the particular that whole areas of possible further investigation and analysis are revealed.

The results applicable to correspondence from Head Office to four of the seven foreign companies studied will be shown or discussed. Companies A and B are foreign subsidiaries which have been trading for many years, the great majority of whose management personnel are local nationals. Companies X and Y are foreign subsidiaries established recently from 'scratch' by the parent company, whose senior management are mostly expatriate personnel.

1. *Annual volumes of correspondence* (Table 1)

This Table shows the overall volume of communication between Head Office and the four selected subsidiaries. It cannot be interpreted without reference to the subsequent Tables which show further detail.

TABLE 1. VOLUMES OF CORRESPONDENCE FROM HEAD OFFICE TO FOUR SUBSIDIARIES

	Subsidiaries				
	A	B	X	Y	Total
1968/69	528	519	341	267	1655
1969/70	467	477	254	302	1500
1970/71	678	664	371	386	2099
1971/72	476	327	642	484	1930

2. *Annual volumes and proportions of mandatory orders in correspondence* (Tables 2 and 3).

The official view of control taken at Head Office is one of decentralisation. Foreign subsidiaries were left largely to run their own affairs. If this was the case in practice, one would expect the volume of items shown in the 'Total' column of Table 2 to be maintained at a constant level. The actual number of mandatory items would be irrelevant.

TABLE 2. VOLUME OF MANDATORY ORDERS IN CORRESPONDENCE FROM HEAD OFFICE

	Subsidiaries				
	A	B	X	Y	Total
1968/69	87	71	110	99	367
1969/70	42	42	65	56	205
1970/71	55	35	29	67	184
1971/72	30	13	45	87	175

TABLE 3. PROPORTION OF MANDATORY ORDERS IN CORRESPONDENCE FROM HEAD OFFICE

	Subsidiaries				
	A	B	X	Y	Total
1968/69	17%	14%	32%	37%	22%
1969/70	9%	9%	26%	19%	13%
1970/71	8%	5.5%	8%	17%	9%
1971/72	6.5%	4%	7%	18%	9%

In fact, the volume of these items is seen to be declining. This fact, by itself, would suggest that the group is in the process of marginally decentralising. However, the total volumes of correspondence in Table 1 are seen to be increasing. It is then the proportion of mandatory orders to total correspondence which will provide the more illuminating interpretation of this data.

Moreover, it is clear from Tables 2 and 1, that the overall trends are not

Vol. 4
1976

consistently reflected in each of the subsidiaries. Only in Company B is this the case. Based on volumes alone, therefore, it would seem that if the group as a whole appears to be centralising or decentralising, the reverse process may be taking place with regard to individual subsidiaries.

Table 3 shows the data in Table 2 expressed in percentage proportions of the totals shown in Table 1. The patterns now emerging can be interpreted with more confidence. The aggregate proportions confirm the existence of a marked process of decentralisation. They also show, with the single but insignificant exception of Company Y in 1971/72, that the four individual companies broadly reflect the overall trend. It was an interesting feature of this study, and somewhat surprising, that *a company could maintain a standardised attitude for all its subsidiaries, irrespective of their histories and indeed of the parent company's varying degree of ownership*. The overall effect need not necessarily be that perceived or intended, but it was constant in direction—all were decentralising. The differences are revealed in terms of the starting points of each and the speed of the process. Table 3 shows clearly that Companies A and B, with their longer histories, had moved further along the process than had the two 'scratch' companies.

3. *Perceptions* (Tables 4 and 5)

Our analysis has so far ignored the people involved in the system. Both Head Office and subsidiary personnel were asked to give their perceptions of what had been happening between their respective organisations in the general terms discussed above. Head Office people consistently underestimated the proportion of mandatory orders sent (Table 4), and in general so too did the subsidiary personnel.[3] Expatriate personnel were more conscious of the degree of control exercised on them than were local nationals in the subsidiaries, but their perceptions were not markedly more accurate.

The implications of these perceptions are interesting. Firstly, both Head Office and subsidiary personnel had no real idea of the extent of control on an organisational basis; both groups of respondents were closely following the stated, though not necessarily actual, group philosophy of decentralisation. Both groups believed that control was minimal. This suggests the possibility that a group of individuals as an organisation may influence or control another group or organisation without either being aware of the extent to which they are doing this.

These perceptions are an interesting comment on this aspect of the overall system as viewed by individual contributors. They do, incidentally, illustrate for the researcher the danger of the faithful use of perceptions where no documentary

[3] In fact this statement needs some qualification. The perceptions mentioned here are specifically in terms of mandatory orders in correspondence. Subsidiary personnel were generally very conscious of Head Office influence emanating from 'involvement' or 'interference', discussed later in the text.

TABLE 4. ACTUAL PERCENTAGE PROPORTIONS OF ALL CORRESPONDENCE CATEGORIES SENT TO SUBSIDIARIES IN 1969/70 COMPARED WITH THE RESPECTIVE MANAGER'S PERCEPTIONS AT TWO OF THE SEVEN MANAGEMENT LEVELS AT HEAD OFFICE

	Subsidiaries					
	A			B		
	Perceived levels		Actual	Perceived levels		Actual
1969/70	1	2		1	2	
Mandatory orders	10	5	17	5	1	14
Suggestions	5	5	6	10	5	6
Advice	12	10	25	Nil	10	37
Queries	45	15	3	25	25	3
Apologies	2	10	3	2	1	4
Information	10	10	12	38	5	14
Requests	5	15	12	5	50	9
Complaints	1	5	1	5	Nil	1
Others	10	25	21	10	3	12
	X			Y		
	Perceived levels		Actual	Perceived levels		Actual
	1	2		1	2	
Mandatory orders	20	10	32	5	10	37
Suggestions	10	10	5	15	10	6
Advice	5	15	24	5	10	27
Queries	48	15	6	20	25	3
Apologies	1	5	1	2	1	1
Information	5	5	6	33	10	9
Requests	5	10	12	5	30	9
Complaints	1	10	3	5	1	1
Others	5	20	11	10	3	7

Vol. 4
1976

evidence is available or accessible. Moreover, this situation suggests that a cardinal rule when taking perceptions should be to rely only on those perceptions which the researcher knows the interviewee is competent to give. In other words the perceptions asked of the interviewee should not place him in the position where he has to extend his own direct experience to what he imagines is that of others, nor where he has to draw heavily on his impressions of what he believes should be happening in an area in which he has minimal involvement, or none at all. By asking the interviewees in this instance for perceptions of their organisations as a whole, the researcher would have been guilty of both these errors, if the intended use of these perceptions had been to understand by this means alone the reality of the situation under discussion.[4]

The researcher then asked interviewees for perceptions of their own contributions to the process. Table 5 shows that there is still a discrepancy between what individuals think they send out and the reality. In this case, the interviewees are

[4] It was mentioned earlier that the researcher's use of these particular perceptions was to locate the company's control philosophy and the commitment of the executives involved in it.

TABLE 5. LEVEL 2 INDIVIDUAL'S PERCEPTIONS OF PROPORTIONS OF MANDATORY ORDERS, SUGGESTIONS AND ADVICE IN OWN CORRESPONDENCE COMPARED WITH ACTUAL

	Subsidiary			
	A		B	
	Perceived	Actual	Perceived	Actual
Mandatory orders				
1968/69		10		10
1969/70	5	9	0	7
1970/71		10		4
1971/72	5	11	20	7
Suggestions				
1968/69		0		5
1969/70	3	4	5	0
1970/71		5		4
1971/72	10	6	10	5
Advice				
1968/69		32		47
1969/70	5	30	5	34
1970/71		11		15
1971/72	15	19	15	9

	X		Y	
	Perceived	Actual	Perceived	Actual
Mandatory orders				
1968/69		25		30
1969/70	5	27	0	19
1970/71		7		13
1971/72	40	7	25	18
Suggestions				
1968/69		0		6
1969/70	10	5	5	2
1970/71		5		4
1971/72	10	12	10	10
Advice				
1968/69		19		33
1969/70	15	30	10	29
1970/71		26		8
1971/72	15	25	15	14

not required to use their imaginations; they are asked to quantify their own direct experience in terms of percentage proportions. They were unable to do this with consistent accuracy, and this underlines the researcher's difficulties in the use of perceptions.

Little wonder that researchers find that Head Office executives talk about 'decentralising' and that in the subsidiary the same system is thought to be

highly centralised. No wonder, too, that subsidiaries often demand a clearer statement of their responsibility and less interference. Tables 4 and 5 explain marked discrepancies in evidence which have often been commented on [3].

The majority of changing figures for perceptions in level 2 in 1971/72 over 1969/70 can be taken to mean no more than a realisation of the facts about what was happening by the individuals interviewed. This is an example of the influence of the researcher on the situation studied.

4. *Control and involvement*

It is clear that if an international group is in the process of decentralising, Head Office is not losing its interest in its foreign subsidiaries. If the decision to decentralise had been implemented to the letter, the subsidiary's gradual autonomy would be reflected in the correspondence analysis in the following manner. The overall trend in the proportion of mandatory orders would decline, and the balance would be distributed more or less equally throughout the other categories. In other words, if all other factors are held constant and mandatory orders are reduced by 8 per cent each year, we would expect to see the other eight categories increase by about 1 per cent each for the same period to compensate for the loss.

If, however, when decentralising Head Office is merely replacing its mandatory control by less obvious forms of 'involvement', we could say that Head Office was maintaining its level of interest, certainly, and of influence, possibly. In this case, we would adjust our interpretation of the speed and extent of decentralisation accordingly.

Evidence to support or deny these arguments lies in the 'involvement' categories of the analysis. Table 6 shows these involvement categories by distinguishing between what has been, and what has not been, requested by the subsidiary. If the advice and information requested were to increase by markedly more than the equivalent of our theoretical 1 per cent per annum, we might interpret this as the subsidiary's need for Head Office influence or guidance. If a similar variance were shown in the categories not requested by the subsidiary, this would suggest Head Office's desire for whatever motive to maintain more or less the influence *status quo*.

Table 6 shows mixed results for involvement requested by the subsidiaries. A and Y have fluctuated over the four years, with no change overall. B is reducing while X is increasing. Further evidence from other sources of data is required before commenting with confidence on this data. Involvement not requested by the subsidiary shows consistent and, in certain cases, marked increases over the four years for all the companies. This would suggest that Head Office's influence is not declining at quite the rate indicated by the reduction in mandatory orders. Indeed, in subsidiaries X and Y overall Head Office influence may not be declining at all.

Vol. 4
1976

TABLE 6. PERCENTAGE OF NON-MANDATORY 'INVOLVEMENT' IN SUBSIDIARY AFFAIRS

	Subsidiaries A				B			
Year	1	2	3	4	1	2	3	4
Requested by subsidiary	25	28	25	23	37	31	33	22
Not requested by subsidiary	21	20	30	41	18	20	24	31

	Subsidiaries X				Y			
Requested by subsidiary	24	24	35	34	27	31	21	27
Not requested by subsidiary	23	21	37	39	18	22	34	36

Note: Year 1 1968/69
2 1969/70
3 1970/71
4 1971/72

5. *Control direction and saturation* (Table 7)

The discussion of control and involvement provides some clarification for the researcher when he is faced with contradictory statements about the state of decentralisation from Head Office and subsidiary interviews. Subsidiary awareness of, and apparently sensitivity towards, Head Office 'interference' is quickly noted by the researcher. It masks a complex web of relationships to which the interviewer by himself has limited access.

By 'direction' we mean the categories of correspondence sent from one management level or function at Head Office to one in the subsidiary. By 'saturation' we mean the aggregate spread of individual relationships between the different levels at Head Office and the subsidiary.

Table 7 is a summarised extract from our attempt to measure direction and saturation. For the two subsidiaries shown, all correspondence from each of the Head Office individuals (shown across the page as 2, 3, 4, 5, 6, 7) concerning budgets, financial and statistical returns and general management items are recorded against eight subsidiary levels (shown down the page as 1, 2, 3, 4, 5, 6, 7, 8). The researcher can now see who at Head Office sent what to whom at the subsidiary. For instance, level 3 at Head Office sent 17 items concerning Financial and Statistical Returns to level 3 in Company B. This information tells the researcher precisely the background from which his interview respondents at both Head Office and subsidiary are talking, and their remarks can be interpreted in the light of this data.

Table 7 shows clearly the marked difference in saturation in subsidiaries B and X. In the former's case, the spread of interaction is almost comprehensive

TABLE 7. THE DIRECTION AND SPREAD OF INDIVIDUAL RELATIONSHIPS BETWEEN HEAD OFFICE AND TWO SUBSIDIARIES—CORRESPONDENCE VOLUMES 1970/71

	Budgets						Financial and statistical returns						General management					
H.O.	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
Sub.								Subsidiary B										
1		2						1					50	19	18		5	124
2		3	1					2	1				20	18	41			1
3		34	2				1	17	7				4	27	30			
4		2	2					2	5				5	11	57			1
5													7	1	3		40	8
6																		
7													2	1	1		1	38
8														1	47	3		
								Subsidiary X										
1		15	5				6	14	23				54	67	136		11	26
2																		
3		6																
4															6			
5														1				
6																		
7																		3
8																		

The numbers across the top and down the side of the Table refer to the level of management in Head Office and in the foreign subsidiary, respectively. The numbers in the squares show how many communications actually pass.

and we would describe this communication pattern on a 'saturation' continuum as multi-directional, whereas the pattern with subsidiary X would be termed mono-directional. In this case, Head Office communicates almost exclusively with subsidiary level 1.

The analysis from which Table 7 is extracted shows for each figure given here a breakdown into the different categories. Hence we can assemble quantified data for control and involvement direction and saturation in the same way that the total of all categories is described here.

Control saturation is an interesting window from which to view communication behaviour.

6. *Communication and control patterns*

We have shown how an analysis of this type can produce data which varies from the very general to the very particular. We have drawn various conclusions from the data shown to illustrate the applications of this research technique. It is, however, but one source of data. In our progression towards the classifica-

Vol. 4
1976

tion of communication and control patterns the conclusions drawn in isolation from this analysis will be qualified by the contributions of other investigations. This analysis provides a useful contribution to our understanding of the complex issues in the centralisation-decentralisation cycle with particular emphasis on the extent to which this process is a function of the nature of the personal relationships involved.

We have briefly described the results to be gleaned from this analysis. It can be adapted to accommodate a multitude of research needs. It also has possible practical applications for the businessman. The analysis was used, for example, by the company in question to show the extent of delegation within the international division. It could be used by a subsidiary to prove its claims of unnecessary interference. Equally, it could be used by an Head Office to reject or qualify such a claim.

For the researcher, perhaps the most important feature of the present method is that it provides a corrective to evidence which has often been accepted and to the resultant conclusions. The company studied *does* have a standard control system, it is *not* a decentralised organisation in the sense in which its executives previously thought it to be.

From this, it follows that the successful performance of the company does not derive from the decentralisation or lack of standardisation as was claimed. The method outlined allows both the characteristics and the claims to be examined and understood.

REFERENCES

1. TANNENBAUM ARNOLD S (1966) *The Social Psychology of the Work Organisation.* Tavistock, London.
2. BUDD RW, THORP RK, DONOHEW, L (1967) *Content Analysis of Communications.* Macmillan, New York.
3. BROOKE MZ and REMMER HL (1970) *The Strategy of Multinational Enterprise.* Longman.

ADDRESS FOR CORRESPONDENCE: *Dr. MZ Brooke, Senior Lecturer in Management Studies, UMIST, PO Box 88, Manchester M60 1QD.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Managerial Authority in a Decision Analysis Context

MH BERHOLD

(*Received September 1974, revised from September 1975*)

The paper uses a decision analysis context to develop a definition of the *managerial* concept of authority. This development begins with a decision analysis model to identify a typical decision problem. This model is then used to develop a structured definition of authority by building on a conceptual definition that was abstracted from the management literature: authority is the power to make action choices. As developed in the decision analysis context, authority is defined as the power to alter resource allocation, incentives or perception. Various statements in the literature are examined to see if the decision analysis model is incompatible with the descriptive model presented by the literature. The paper then briefly discusses possible uses of the model in analyzing authority delegation, the perceived authority concept, and the relationship between authority and responsibility.

THIS PAPER uses a decision analysis context to develop a structured definition of the managerial concept of authority. This definition provides the basis for formalizing what is currently *known* about the managerial concept of authority and thus provides a better understanding of that concept, which is defined in the management literature simply as: authority is the power to make action choices. The definition adopted in this paper is consistent with that presented in the literature. This initial test of consistency with the current literature is necessary because, although we are developing a definition of authority within a normative framework, we must be reasonably sure that the normative model is a reasonable approximation of, or at worst a subset of the descriptive model that we find in that literature. If this is not done, then our normative model will have no apparent connection with the descriptive model.

It might be useful to reinforce a particular viewpoint regarding one role of normative models that is often underrated. As stated by Aaker [1, p. 593], "The benefit of a normative model is usually not the machinery it makes available but rather, the vocabulary it creates, the structure it provides, and its suggestions for future research." This is the main thrust of this paper.

In discussing the model some future research topics are considered. These topics involve the concepts of perceived authority, authority delegation, and the authority–responsibility relationship. This gives an indication of areas where the model might be fruitfully applied.

I. THE DEVELOPMENT OF A DEFINITION OF AUTHORITY

Basic decision model

A decision maker's basic problem is to choose among the feasible action choices that he has identified. Consider the actions of the decision maker in a physical system such as a car. The decision maker (driver) can choose various settings for the steering wheel, brake pedal, gas pedal, gear shift, etc. These actions have specific restrictions that cannot be violated. For example the steering wheel can be turned only so far, the gearshift has certain positions. All of these feasible actions can be categorized under one heading—the feasible set of actions involved in driving a car.

The outcome for the decision maker can be a function of incentives (positive or negative) that are contingent on some measure of performance [3]. For example an overdrawn checking account, speeding in a car, trespassing, etc. all exact penalties related to the degree of violation.

To further illustrate: a department manager is authorized to purchase equipment from outside the company by using a purchase order that requires only his signature. However, there is a statement about the upper limit on the total purchases for the year; consequently, while his feasible actions have been increased by the authorization to make purchases, his total purchases are controlled in the sense that he is told not to exceed the upper limit. He will not exceed the upper limit because of the possible penalties (negative incentives). The incentive aspect can be removed by requiring another signature that could be obtained only if the total purchases for the year do not exceed the upper limit.

Vol. 4
1976

In our example there is an option between restricting actions and introducing incentives, with economics likely dictating the choice. However, when restrictions cannot be introduced, then we have to rely on a performance characteristic that we can relate an incentive to. For example we cannot observe a lathe operator to determine if he runs the lathe at the proper speed, however, it may be possible to observe how often the cutting tool must be sharpened. Thus, we could relate an incentive payment to the frequency of sharpening.

The decision maker's problem can be summarized in more abstract terms as follows

$$\text{Max}_{x \in X}\, v[x/P(\alpha)] \tag{1}$$

where v is the decision maker's value system such as expected value, expected utility, minimax, etc. $X = \{x\}$ is the set of feasible actions and $P(\alpha)$ is his perception of those aspects that are relevant to the given value system with α indexing the incentive system. This representation of the decision maker's problem will be used to develop a definition of authority.

One point merits additional emphasis. The model explicitly involves the relevant perceptions of the decision maker. Thus, in discussing an individual's authority, we are actually concerned with his perceived authority. So, unless specified otherwise, the term authority will henceforth mean perceived authority.

The incentive concept, as we use it, involves two persons: one, denoted as the principal, is the decision maker who provides the incentive, and the other, denoted as the agent, is the decision maker receiving the incentive payment. These roles of the principal and agent are important in our subsequent definition of authority.

Authority defined in the context of the basic model

We begin with a conceptual definition of authority as abstracted from the management theory literature. To restate it: a person's authority is, simply, the power to make action choices.

Paraphrasing this in the context of our previous development we say that the principal's authority with respect to the agent is viewed as the power to influence the action choice of the agent. We will call this indirect authority since the principal is exerting control indirectly through the agent. This is contrasted with direct authority which is the principal's direct control of his own actions. The following development will focus on indirect authority.

From (1) we see that the optimal decision of the agent can be influenced by the principal through his choice of an incentive function or his control over the feasible set X. For purposes of identification, we will use the term 'resource allocation', to refer to the principal's control over the agent's feasible set of actions.

To summarize, we can identify two distinct types of authority that the principle may have with respect to the agent: control of the incentives provided to the agent and control of the resources provided to the agent. The reader may have observed that in a management theory context we might call the principal a manager and the agent a subordinate.

The basic two-person model can be extended by introducing another decision maker who controls the principal's authority by specifying the incentives that can be made available and what resources can be allocated. Hierarchical structures of this nature are common to most discussions of authority. We might further note that this binary, two-person, model can be the basic building block of a more comprehensive structure such as the theory of teams [14].

With this in mind we will next evaluate authority, as defined in this paper, to determine its equivalence to authority as defined in the management theory literature. For identification purposes only, the concept of authority as developed in this paper will be denoted as the *authority model.*

II. THE MODEL'S CONSISTENCY WITH THE LITERATURE

Before beginning, it should be mentioned that this survey is not meant to be exhaustive, merely representative. The survey is adequate as a preliminary test to the next testing process, that being an appearance in the literature and its subsequent test through application.

As an initial objective, a literature search would seek to uncover statements that directly support the authority model. However, we quickly discovered that the literature defines authority largely in a descriptive sense rather than within the context of a specific structure such as the authority model. Thus, the search changed its objective and focused on statements in the literature that might contradict the model.

Management theory literature

As abstracted from the management theory literature: a principal's authority is the power to influence an agent's action choices. The following statements support this definition. ". . . the standard definition is 'legal or rightful power, a right to command or to act.' " [10, p. 59]. "Authority is the right to give orders and the power to exact obedience" [5, p. 21]. "Authority can be defined as the capacity to evoke compliance in others [15, p. 86]. "Applied to the managerial job, authority is the power to command others, to act or not to act in a manner deemed by the possessor of the authority to further enterprise or departmental purpose" [10, p. 59].

Koontz and O'Donnell discuss two theories of authority—the formal authority theory and the acceptance theory and [10, p. 61] use Barnard [2, pp. 162 ff] for the definition of formal authority: "The concept of authority as being power transmitted from basic social institutions to individual managers has been called formal authority." Commenting on this, Koontz and O'Donnell state, "Most formal authority theorists emphasize the legal aspects of private property as the source of authority, though good sociological analysis would broaden the source to include all related social institutions."

A contrast to the formal theory is the acceptance theory defined by Barnard [2, p. 163]: "Authority is the character of a communication (order) in a formal organization by virtue of which it is accepted by a contributor to or 'member' of the organization as governing the action he contributes; that is, as governing or determining what he does or is not to do so far as the organization is concerned. According to this definition, authority involves two aspects: first, the subjective, the person, the accepting of a communication as authoritative . . .; and second, the objective aspect—the character in the communication by virtue of which it is accepted."

The authority model tends to confirm the formal authority theory. The principal is in a position to actually dictate the incentive function, and the

Vol. 4
1976

resource allocation, although he rarely does. When we say the principal dictates the incentive function, we do not mean that the agent must accept it. He can take advantage of other opportunities. However, the point remains: the principal can dictate the incentive function.

Koontz and O'Donnell mention one other theory and that is the competence theory. In their words: "Under this heading is the individual who has made, in effect, subordinates of others through the sheer force of personality and the engineer or economist who exerts influence by furnishing answers or sound advice" [10, p. 63]. This theory introduces another classification to our authority model. Anyone who can alter an individual's perception can have an influence on that individual's decision. With this in mind, let us augment the authority model accordingly.

The authority model is useful in testing various definitions. For example consider Simon's statement in regard to a general theory of authority: "The power in charge of unifying common action through rules binding for us all is what everyone calls authority" [16, p. 48]. This statement refers almost solely to the incentives. So by *rules binding for us all*, Simon seems to mean uniform incentives (positive or negative) which are applied to us all. Since Simon's so-called general definition does not consider resource allocation, or competence, we may conclude that his definition is not complete.

Literature outside the area of management theory

The concept of authority is an important concept in other social sciences. Thus it might be useful to identify some connecting links with literature in other fields. However, we should emphasize that our primary focus is directed towards the management context.

The fields of sociology and political science contain a line of research that can be traced to Goldhamer and Shils [7]. Their discussion begins with a conceptual definition: A person has power to the extent that he ". . . influences the behavior of others in accordance with his own intentions" [7, p. 171]. They further identify power in terms of three types [7, pp. 171–172]: (1) Force in terms of physical power, (2) Domination in terms of an agent letting his goals be specified by a principal, and (3) Manipulation of the agent's alternatives or his perceptions of the decision making environment (beliefs about cause and effect relationships or actions that might be taken by other decision makers).

In this same vein Dahl [4, p. 203] states, ". . . A has power over B to the extent that he can get B to do something that B would not otherwise do." And French [6, p. 183] amplifies this by identifying five types of power within that definition:

(1) Attractive	B's liking for A
(2) Expert	B's perception that A has superior knowledge and information.
(3) Reward	A's ability to mediate rewards for B.

(4) Coercive	A's ability to mediate punishment for B.
(5) Legitimate	B's belief that A has a right to prescribe B's behavior or opinion.

These five types of power can each be related to the authority model: Attractive power would involve preferences on outcomes. Expert power involves perception. Both Reward and Coercive power would be included under incentives and Legitimate power would involve the principal prescribing the agent's beliefs, values or actions.

Harsanyi [8] and March [12], [13] move away from the problem of definition to the problem of measurement. Harsanyi identifies the following five aspects that are relevant in measuring the principal's power.

(1) Scope	Actions principal can get agent to do.
(2) Base	Resources available to principal.
(3) Means	Specific actions principal can use to influence agent.
(4) Amount	Probability of agent's response.
(5) Extension	Number of agents controlled by principal.

In line with this he discusses two kinds of costs. One is the principal's cost of exercising control over the agent. The other is the agent's cost if he refused the principal's control.

Another related line of research is found in the social psychology literature and is concerned with the concept of "dyadic interaction". Dyadic Interaction (DI) involves two persons, each of whose behavior is not independent of the previous behavior of the other person.[1] It is worth noting that the DI representation uses the social interaction context as its primary referent, a referent that is somewhat different than the formal organization type of context that serves as a referent for the authority model.

The DI concept is built up from a two person model (as dyadic implies) and involves a sequence of actions and responses where each person's responses are contingent on previous actions (of both). This idea of contingent responses is then used to consider the major ways in which one person's responses become stimuli and thus control another person. Two kinds of control are identified—outcome control and cue control.

Outcome control is similar to control of incentives or to the related concepts of reward and punishment. On the other hand, cue control seems to go beyond the authority model. Cue control appears to involve two facets, one being control through the provision of information and another is control through the provision of cues that can predictably elicit desired responses. The control through information is, in the authority model, the power to alter perception.

On the other hand the provision of cues involves stimuli that elicit pre-established habit patterns by another person in a manner similar to conditioned

[1] See [9, pp. 505–536] for a more extensive discussion.

responses. Managers are almost surely aware of this facet when they deal with subordinates; however, they might not consider the extent of its significance.

As discussed by Jones and Gerard [9, pp. 537–590], much of the experimental literature relating to the DI concept is, interestingly enough, concerned with the two-person zero and nonzero-sum games that were developed with game theory [17]. Much of the experimental work involving the DI model considers the very interesting behavioral problem that is contained in the nonzero-sum game called the prisoners dilemma [11, pp. 95–102]. This reliance on game theory by the DI model (as well as Harsanyi [8]) illustrates the useful role that normative models often serve in empirical research, in terms of providing structure as well as concepts.

This brief review of the DI concept indicates that there is a significant relationship between that concept and the authority model. This follows from the fact that the DI concept, like the authority model, deals with the idea of one person exerting control over another. Moreover, the DI structure offers a great deal of generality because it is defined in a largely context-free manner and so, it, in no sense, excludes our principal–agent (two person) authority model.

We further note that the DI concept has an extensive literature on experimental results. So, given the structural compatibility with the authority model, it would seem that a good deal of these results can be related to the authority model. So the DI concept is of added importance because its experimental literature represents a rich potential for adding to our knowledge of authority relationships.

III. SUMMARY AND DISCUSSION

We have, using a two-person model, described three distinct types of indirect authority: (1) control of incentives, (2) control of resources, and (3) influence of perception. These are confirmed in three diverse bodies of literature, although no single author embraces them all.

At this point we have hopefully provided a basis for the development of a much more elaborate and extensive theory upon which we can establish a stronger relationship between management theory and decision analysis. In the many areas where management theory and decision analysis refer to a problem of common interest, one expects to find a mutuality of interests that would give rise to an interdisciplinary effort. This effort would offer the promise of substantial benefits, since it would combine the extensive data from management theory with the analytical techniques of decision analysis.

However, some ingredients, essential to an interdisciplinary effort, are missing. A common vocabulary does not exist. There is not even a *common* reference in terms of a basic model or models. Many terms in management theory are largely conceptual; many terms in decision analysis lack good references within the manager's environment.

As an indication of possible applications of the model, it might be useful to briefly discuss three concepts that could fruitfully use the authority model for future research. These concepts involve perceived authority, authority delegation, and the authority–responsibility relationship.

Perceived authority is an individual's authority as perceived by himself or by others. Thus a person perceives his own authority as well as others perceiving what his authority might be. These various perceptions may be inconsistent with each other as well as being inconsistent with reality. It would seem that the structure afforded by the authority model would be of value in discussing the concept of perceived authority, particularly as it might be considered in terms of perceptions being updated with the receipt of information.

Authority delegation could discuss the delegation of each type of authority: control of resources, incentives and information channels. The last type would be somewhat complex because the principal would have to be in a position to change an agent's perceived value as an information source.

The final suggested direction of research would involve a consideration of the authority–responsibility relationship as it might be viewed in the authority model. At a simplistic level we can view authority as the power to exercise control over future events either through direct control or through indirect control. In this same light we can view responsibility as an evaluation of the resulting events *vis à vis* the control that existed. The authority model, with its decision analysis reference would be very useful in discussing the cause and effect relationship that is implied by the authority–responsibility relationship.

Vol. 4
1976

These lines of research appear to be fruitful directions because the authority model provides a more structured reference. This structure reduces the haziness associated with the concept of authority and thus provides the basis for improving the educational and training processes in both academic and nonacademic environments. Consequently, it is possible to communicate and to understand the basic concept of authority in terms of specific characteristics as opposed to the hazier inferential learning process.

We should be continually reminded that history provides us with numerous examples of formal models providing the basis for sharpening hazy concepts. For example, in the distant past the term 'expectation' was initially a somewhat hazy concept. This concept has, in many contexts, now been given a much sharper definition because of the expected value operator that has been defined in probability theory. In line with this, we have also seen formal definitions of probability sharpening our once hazy concepts regarding uncertainty. More recently the idea of subjective probability has further sharpened these ideas of uncertainty.

As hazy concepts become defined in more specific terms, then we are able to use them more effectively in communicating with one another. In addition, we are able to use these concepts more effectively in our research and analysis.

Thus, in attempting to formalize various existing concepts, we must choose a middle ground somewhere between the mathematician's requirement that the model have a sophisticated internal logical consistency, and the empiricist's requirement that the model be a 'best' fit to what has been observed. This paper has hopefully developed a model that has adjusted itself to these requirements of external validity and a sophisticated internal consistency.

REFERENCES

1. AAKER D (1973) Toward a normative model of promotional decision making. *Mgmt Sci.* **19** (6), 593–603.
2. BARNARD CI (1960) *The Functions of the Executive*. Harvard University Press, Cambridge.
3. BERHOLD M (1971) A theory of linear profit sharing incentives. *Q. J. Econ.* **85** (3), 460–482.
4. DAHL R (1957) The concept of power. *Behavl Sci.* **2**, 201–215.
5. FAYOL H (1949) *General and Industrial Management*, Trans. from French (Dunod) by Constance Storrs, Sir Isaac Pitman, London.
6. FRENCH J (1956) A formal theory of social power. *Psychol. Rev.* **63** (3), 181–194.
7. GOLDHAMER H & SHILS JE (1939) Types of power and status. *Am. J. Sociology*, **45** (2), 171–182.
8. HARSANYI J (1962) Measurement of social power, opportunity costs and the theory of two-person games. *Behavl Sci* **7**, 67–80.
9. JONES E & GERARD H (1967) *Foundations of Social Psychology*. Wiley, New York.
10. KOONTZ & O'DONNELL C (1968) *Principles of Management*. Fourth Ed. McGraw-Hill, New York.
11. LUCE R & RAIFFA H (1957) *Games and Decisions*. Wiley, New York.
12. MARCH J (1955) An introduction to the theory and measurement of influence. *Am. Polit Sci. Rev.* **49** (2), 431–451.
13. MARCH J (1957) Measurement concepts in the theory of influence. *J. Politics*, **19** (2), 202–226.
14. MARSCHAK J & RADNER R (1972) *Economic Theory of Teams*. Yale University Press, New Haven.
15. PRESTHUS R (1960) Authority in organizations. *Publ. Adm. Rev.* **20** (2), 86–91.
16. SIMON Y (1962) *A General Theory of Authority*. University of Notre Dame Press, Notre Dame.
17. VON NEUMANN J & MORGENSTERN O (1947) *Theory of Games and Economic Behavior*, 2nd Ed. Princeton University Press, Princeton.

ADDRESS FOR CORRESPONDENCE: *Professor MH Berhold, Georgia State University, 33 Gilmer Street, S. E. Atlanta, Georgia 30303 USA.*

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain.

The Modelling of Supply in the Shipping Industry

AJ TAYLOR

(*Received June 1975; in revised form October 1975*)

This paper considers the way in which certain variables in the shipping industry fluctuate. Factors such as supply and demand for cargo carrying capacity, freight rates, laid-up tonnage, are shown to form a number of interacting feedback loops, and it is demonstrated that it is possible to construct a dynamic model of the system.

INTRODUCTION

WITHIN THE shipping industry there exists a great deal of fluctuation in various factors such as freight rates and the amount of unemployed and laid-up shipping. Such fluctuations can be extremely important in the operations of shipping and chartering companies. Consider, for example, the situation at the present time in the area of oil-tanker employment. Demand for oil tankers is low, freight rates are low and so the amount of tonnage being laid-up is escalating. However, to lay-up a ship is not to reduce costs to zero because certain costs are, necessarily, still incurred. A recently quoted figure [1] for the costs in lay-up for a very large crude carrier (VLCC) is $2300–$2600 per day, or almost $1M per year. If a company could know what combinations of market conditions would lead to such large sweeps in important variables then some planning in advance would be possible. A number of authors [2, 3 for example] have examined certain relationships in the market for shipping services from the static point of view. No dynamic analysis has been attempted in this field and it is the purpose of this paper to describe the approach being used by the author in the construction of a dynamic model.

AN APPROACH TO THE MODELLING OF SUPPLY

On examination the shipping industry, or that part of it concerned with the transportation of goods within the international shipping markets, comprises a number of interacting feedback loops.

The basic loop in such systems is a negative feedback loop which may be described as a self-regulating, or goal-seeking, mechanism. Such a mechanism consists of a 'level' variable, (L), such as the number of ships in service at any time, and a 'rate' variable (R), such as the rate of delivery of new ships over a particular time period.

These variables, L and R are related in a manner which may be described in the following way:

$$L_2 = L_1 + \int_1^2 R dt$$

$$R = f(L, \text{and other factors}),$$

where the subscripts 1 and 2 relate to two consecutive points in time. This means that the input rate R depends upon L, which in turn depends upon R. It will be seen that an increase in R leads to an increase in L which, in a negative feedback loops, subsequently has the effect of decreasing the rate R. Hence the self regulation inherent in such a mechanism. In a managed system the input rate R could be a decision—or a policy—for the management of the system. Complications exist in real systems due to problems of identifying the 'other factors' and in the delayed effect of changes in R being felt by L. How can this be applied to the problem of modelling a sector or sectors of the shipping industry? Firstly by identifying the various circular cause–effect relationship which exists. As an example of such circular causality consider the following extract from the report of a recent commission on shipping [4]. "In the longer term the course of freight rates has (also) been affected by bursts of shipbuilding orders placed in response to what may only be short term increases in rates".

Vol. 4
1976

The next stage is to examine such relationships in order to provide a quantitative basis for the construction of the state equations. To take a systematic approach to the description of the various feedback processes operating within the industry, let us consider the system to comprise several interdependent sectors: Shipbuild Orders and Deliveries, Scrappings, Lay-Ups, Change of Trade, Freight Rate Formation. Further, in order to be brief, let us describe the modelling approach as applied to only two of these sectors.

SHIPBUILD ORDERS AND DELIVERIES

With regard to the ship owners' decision to order a new vessel there are many factors to be considered, some of which are:

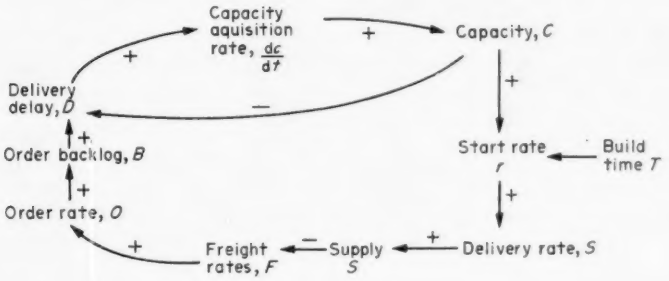
1. Expectations regarding (a) Over-supply in future time-period;
   (b) profitability of particular vessels and markets.

2. Existing prices at shipyards.
3. Backlog of orders at shipyards, and subsequent delivery delay.
4. Possibilities of obtaining finance.

As to the question of which of these factors are most important in determining the size of orders placed at any time consider the data illustrated in Fig. 1.

FIG. 1. *Fluctuations in orders and freight rates (Dry cargo).*

Here there is a clear relationship between the size of orders and the spot charter (single voyage) freight rates. Similar patterns of variation have occurred at other periods of time and, in fact, for the period of Fig. 1 the statistical-



$O = f(F)$
$B = \int Q.dt - \int r.dt$
$D = B/C$
$\frac{dc}{dt} = f(D)$
$s = r$ delayed by D and T
$S = \int s.dt$ − Scrappings, lay-ups, trade changes

FIG. 2. *Relationships in the order–build process.*

correlation coefficient is 0·76 for a lag of 2 months in the freight rate data. The effect of a rise in freight rates is to induce owners to order vessels earlier than they would otherwise have done. It would appear that the most important determinant of the size of orders is the profitability at that time. The interrelationships between some of the factors which help to produce the dynamics of the system are shown in Fig. 2. This sector of the industry has a very low response time to changes in demand. The change in supply of new vessels is not immediate upon the occurrence of changes in demand, and the very large delays inherent in the system help to produce the, sometimes extreme, fluctuations which do occur. This is not to say that there are not other factors which also assist in this process. The policies used, for example, in deciding how much tonnage to order appear to be based upon the level of freight rates at any time and the ship-owners seem to have no memory of past periods when large influxes of orders served to depress the market some years later when the vessels were leaving the shipyards. This kind of situation exists at the present time in the tanker market where there is a gross over-supply of tonnage because of over-ordering in the profitable years of the early 1970's.

CHANGE OF TRADE

The two important trade divisions within the industry are the tanker and dry cargo (bulk carrier) sectors, whose markets are fairly well defined. Uneconomic operation in a particular market, will lead an owner to seek more profitable employment in another market, if this is a feasible proposition. It often happens, in fact, that tankers move into the grain trade when oil-transportation rates are low, and 'combined carriers', having the ability to carry ore, oil and/or other bulk cargo, have a great flexibility in being able to change trades fairly easily.

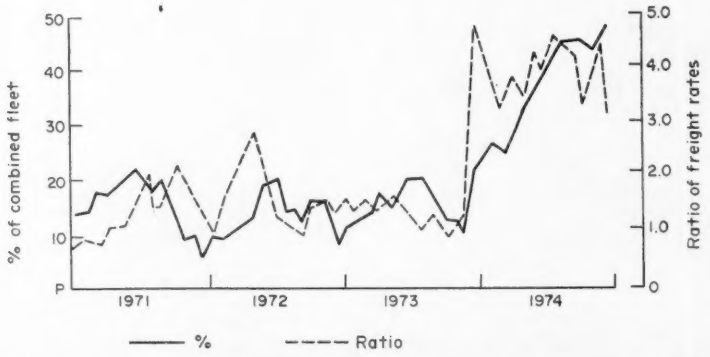


FIG. 3. *Fraction of combined fleet trading in dry cargo compared with the ratio of dry cargo—oil tanker freight rates.*

The relative profitability of one market to another may be measured in terms of the ratio

Dry Cargo Freight Rates/Tanker Freight Rates

(Tanker freight rates are indexed on a scale known as Worldscale.)

Figure 3 illustrates the variation in this ratio over a number of years, together with a comparison of the proportion of the total combined carrier fleet operating in the dry cargo market. A statistical correlation of 0·84 is obtained for a lag of one month in the ratio. Thus, as the freight rates vary the combined-carrier tonnage in the dry cargo sector varies, this having the effect of altering the dry cargo and tanker freight rates. A perpetual fluctuation is set in motion and as long as the industry behaves in the same manner the fluctuation will not cease.

THE MODEL AS A WHOLE

Only two sectors of the model have been described, albeit briefly, but the interdependence of some factors on others is clearly visible. When all the, previously defined, sectors are considered a network of cause–effect relationships is obtained, with the form of the example shown in Fig. 2. These relationships have been analysed and quantified so as to provide the equations for a computer simulation of the behaviour of the dry cargo market. Certain variables are, at this time, exogenous, these being the demand for cargo-carrying capacity and the values of the tanker freight rates. An example of the output of the model is shown in Fig. 4. The demand in this case has been kept constant.

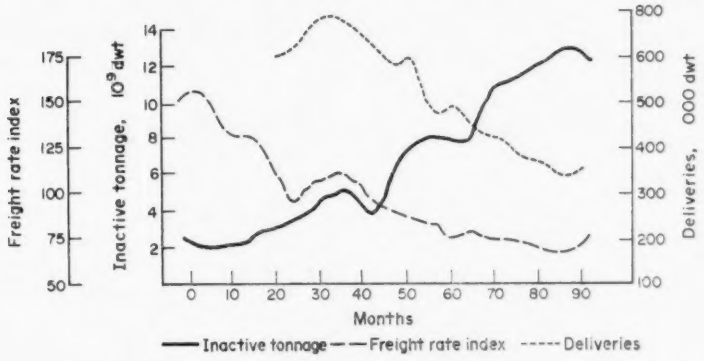


FIG. 4. *An example of model output.*

FUTURE DEVELOPMENTS

The present model is a simplified, but not unrealistic, representation of the real system under study. Developments to the model will include a disaggregation of the variable 'tonnage in service', a more refined approach to the modelling

of various 'delays' in the system, and the treatment of the variable 'laid-up tonnage' to allow it to act as a buffer between ships in service and the scrapyard.

CONCLUSIONS

It has been shown that it is possible to approach the modelling of the dynamics of supply in the shipping industry by use of the concept of feedback. With the assistance of such an approach it should be possible to predict the behaviour of ocean freight rates, quantity of laid-up tonnage, and other factors, given the possibility of alternative trends in shipping demand. Also, it should be possible to provide answers to specific policy questions such as what type of ship will be most profitably employed under certain future market conditions.

REFERENCES

1. *Sunday Times* (1975) 18 May.
2. THORBURN T (1960) *Supply and Demand of Water Transport*. The Business Research Institute at the Stockholm School of Economics, Stockholm.
3. ZANNETOZ Z (1966) *The Theory of Oil Tankship Rates*. M.I.T. Press.
4. HMSO (1970) *Report of the Committee of Inquiry into Shipping*, para. 524.

ADDRESS FOR CORRESPONDENCE: *AJ Taylor, Department of Mathematics, Liverpool Polytechnic, Byrom Street, Liverpool, UK.*

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Stochastic Dominance for Ranking Ventures: Comments and Extensions

JOSEF HADAR

(*Received June 1975*)

This paper suggests some extensions to Tilley and Eilon's paper on ranking ventures. Tilley and Eilon consider concave utility functions consisting of two linear segments and this allows all rectangular distributions to be ordered. Such utility functions have a wide scope and allow other distributions to be ordered.

ol. 4
976

THE DISCUSSION by Tilley and Eilon in their recent paper on the ranking of ventures [5] brings out some of the methodological issues inherent in such problems. One desirable property that an ordering criterion should have is that the criterion should apply to a fairly large group of decision makers. For example, the Second Degree Stochastic Dominance (SSD) criterion applies to the set of all risk averters (all decision makers with concave utility functions); if one venture dominates another by SSD, then all risk averters are unanimous in their preference for the dominating venture, and vice versa. But this equivalence between the SSD condition and unanimous preference by all risk averters is not attained without cost; the SSD condition is such that, given some set of ventures, the condition may not hold between all pairs of ventures in that set. In other words, the SSD condition provides only a partial ordering of ventures.

But such is the case also for other ordering criteria, for example, the mean-variance criterion. If one venture has a higher mean *and* higher variance than another venture, the two ventures cannot be ranked by this criterion. In general, however, the set of distributions that can be ranked by the mean-variance criterion is larger than that which can be ranked by SSD. On the other hand, since the mean-variance criterion applies only to all quadratic utility functions (assuming no restrictions on the admissible probability distributions), the set of decision makers to which it applies is smaller than that for which SSD is the appropriate criterion. There exists, therefore, a trade-off between the size of the set of decision makers to which the criterion applies, and the size of the set of distributions that can be ordered by that particular criterion. In fact,

the only way to achieve a complete ordering of ventures (that is, to reduce the efficient set to a single venture) is to reduce the set of utility functions to one specific function (including all its linear transformations). Then all ventures can be ranked on the basis of expected utility, and the venture that maximizes expected utility is chosen. But problems involving one specific utility function are obviously uninteresting, so that in all the interesting problems involving the ranking of ventures (i.e. two or more utility functions) one should be prepared to accept the inevitable fact that the chosen ordering criterion will induce only a partial ordering of ventures.

Practical considerations may also play a role in the choice of a ranking criterion. From this point of view, criteria involving moments of the distributions (mean, variance, semi-variance, etc.), may have a slight edge over stochastic dominance since for the application of the latter criteria one needs to have information about the entire distribution rather than just about some of its moments. But one may safely expect that the use of high-speed computers in testing for the existence of stochastic dominance will tend to minimize this difference.

Considerations such as the above may have led Tilley and Eilon, and also Flavell [2], to consider a family of concave utility functions consisting of two linear segments. Tilley and Eilon have shown that, given such a family of utility functions, one can order all rectangular distributions. As it turns out, such families of utility functions have a much wider scope than their example may suggest. In particular, one can obtain an ordering criterion for distributions (not only rectangular ones) that is both necessary and sufficient for preference between pairs of distributions for all utility functions belonging to the above class of functions. The formal results concerning this ordering criterion are presented in the following section.

AN ORDERING CRITERION FOR LINEARLY CONCAVE UTILITY FUNCTIONS

We consider concave and continuous utility functions that consist of two linear segments. Such functions will be referred to as *linearly concave*. To avoid any difficulties with regard to the existence of expected utility, we shall assume that the domain on which the utility functions are defined is bounded from below by some point a, but no upper bound is imposed on the domain. Since the decision makers are assumed to satisfy the axioms of expected-utility maximization, the utility functions are unique up to a linear transformation. This permits us to choose scale and origin for each utility function. We, therefore, assume that the slope of the lower line segment of each utility function is one, and that at the point at which the two line segments are joined together, the value of the utility function is zero. These assumptions are formalized in the following definition of a typical utility function:

$$\phi(x) = \begin{cases} x - k \text{ for } a \leq x \leq k, \\ c(x - k) \text{ for } k < x, \end{cases}$$

where $0 \leq c \leq 1$. The point k is the point at which the utility function has a kink.

To facilitate the analysis, we consider a typical subset of all linearly concave functions defined on the above domain. Specifically, a typical subset, denoted by V_z, is the set of all linearly concave functions whose kinks occur at some point between a and z, or at one of these points. Thus, the set V_z defines a two parameter family of utility functions, the two parameters being c and k. (Note that if $k = a$, or $c = 1$, the utility function is simply a straight line.)

As for the probability distributions, for the sake of simplification we consider only distributions with continuous densities and finite means.[1] Densities will be denoted by f and g, and the associated distribution functions by the respective capital letters. Like the utility functions, the probability distributions are defined at all points from point a to infinity.

One new concept that will be introduced is the notion of *semi-SSD with respect to point z*. Denoting this binary relation by S_z, we define the following:

$$F \; S_z \; G \text{ if and only if } \int_a^y [F(x) - G(x)]\,dx \leq 0 \text{ for all } y \text{ satisfying } a \leq y \leq z.$$

Clearly, as z approaches infinity, semi-SSD approaches SSD. It is also obvious that SSD implies semi-SSD; thus, the latter is a weaker condition.

We now prove a general theorem on ranking of distributions that applies to all utility functions in V_z. In this theorem, a symbol such as $\bar{u}_F$ shall denote expected utility under distribution F, and m_F denotes the mean of F; and similar notation for distribution G.

THEOREM 1

$$F \; S_z \; G \text{ and } m_F - m_G \geq 0 \text{ if and only if } \bar{u}_F - \bar{u}_G \geq 0 \text{ for all } \phi(x) \in V_z.$$

In words, if F dominates G in the sense of semi-SSD with respect to z, and the mean of F is at least as large as the mean of G, then all decision makers with utility functions in V_z prefer F to G, or are indifferent between them; and vice versa.

[1] The assumption of finite means, and the lower boundedness of the utility functions, guarantee that expected utility always exists (is finite). See Arrow [1].

PROOF

Consider an arbitrary utility function in V_z. Then the difference in expected utility under distributions F and G is defined as

$$\bar{u}_F - \bar{u}_G = \int_a^k (x-k)[f(x)-g(x)]\,dx + c\int_k^\infty (x-k)[f(x)-g(x)]\,dx. \tag{1}$$

It can be shown that the above equation is equivalent to

$$\bar{u}_F - \bar{u}_G = \int_a^k [G(x)-F(x)]\,dx + c\int_k^\infty [G(x)-F(x)]\,dx. \tag{2}$$

For notational convenience, let us rewrite equation (2) as

$$\bar{u}_F - \bar{u}_G = A + cB, \tag{2'}$$

where A and B denote the respective integrals in equation (2).

First, we prove that the semi-SSD condition together with the restriction on the means implies preference for F over G. Since $A + B = m_F - m_G$, we have $A + B \geq 0$ by the restriction on the two means assumed in the theorem. This implies $B \geq -A$. Substituting for B in equation (2') gives $\bar{u}_F - \bar{u}_G \geq (1-c)A$. But $A \geq 0$ by the semi-SSD assumption, while $1 - c \geq 0$ by the definition of a typical utility function in V_z. Consequently, $\bar{u}_F - \bar{u}_G \geq 0$ for all $\phi(x) \in V_z$.

We now prove the converse, using a proof by contradiction. Thus, we assume that $\bar{u}_F - \bar{u}_G = A + cB \geq 0$ for all $\phi(x) \in V_z$. Suppose now that the conclusion of the theorem is false, that is, assume that for some k, $a < k \leq z$, the semi-SSD condition fails to hold, meaning $A < 0$. But then for $c = 0$ we get $\bar{u}_F - \bar{u}_G < 0$, contradicting the assumption of the theorem. Hence we must have $A \geq 0$.

Now assume that the restriction on the means fails to hold, that is, assume $A + B < 0$, while $A \geq 0$. Then there exists a positive constant e such that $B = -(A + e)$. Substituting for B in equation (2') yields

$$\bar{u}_F - \bar{u}_G = A - c(A + e). \tag{3}$$

Now, pick a utility function in V_z whose slope coefficient satisfies the inequalities $(A/A + e) < c < 1$, which implies $A < c(A + e)$. Substituting for A in equation (3) gives $\bar{u}_F - \bar{u}_G < c(A + e) - c(A + e) = 0$, which again is a contradiction. Hence we conclude that $A + B \geq 0$, which completes the proof.

What the above theorem demonstrates is, first, that there exists an ordering criterion for ordering distributions (not only rectangular distributions) such that whenever one distribution dominates another by that criterion, all decision makers in the corresponding set of linearly concave functions (i.e. V_z) will

unanimously prefer the dominating distribution, and vice versa. In its form, the above theorem is similar to theorems for ranking uncertain prospects that are based on stochastic dominance criteria, and particularly those using the SSD condition.[2] It is, of course, not surprising that, since the set V_z is a subset of all concave functions, the ordering criterion for V_z, that is, semi-SSD, is weaker than SSD. But because of that, the semi-SSD criterion is capable of ordering a larger set of distributions than that which can be ordered by SSD. Another distinction between the two approaches is that semi-SSD is more flexible than SSD, since by varying the point z, one can change the set of utility functions under consideration. For example, when a higher value of z is selected, the set V_z becomes larger, and the corresponding semi-SSD condition becomes stronger. And conversely for decreases in z.

In the limit the two approaches coincide. Thus, when z is made to approach infinity, the subset V_z is extended to encompass *all* linearly concave functions, and the semi-SSD condition is replaced by SSD. In that case, Theorem 1 can be stated as follows:

THEOREM 1′

$F \; S \; G$ if and only if $\tilde{u}_F - \tilde{u}_G \geq 0$ for all $\phi(x) \in V_\infty$.

In the above, the symbol S denotes the SSD relation. The restriction on the two means is omitted since it is implied by the SSD condition. For purposes of comparison, let us restate the basic theorem using SSD (e.g. Proposition 2 in [4]).

THEOREM 2

$F \; S \; G$ if and only if $\tilde{u}_F - \tilde{u}_G \geq 0$ for all $\phi(x) \in U_2$.

In the above theorem, the symbol U_2 denotes the set of all concave (and continuous) utility functions defined at points between a and infinity. Since in both theorems the statements preceding 'if and only if' are the same, it immediately follows that the statements succeeding 'if and only if' are also equivalent. In other words, one can state the following corollary:

COROLLARY

$\tilde{u}_F - \tilde{u}_G \geq 0$ for all $\phi(x) \in V_\infty$ if and only if $\tilde{u}_F - \tilde{u}_G \geq 0$ for all $\phi(x) \in U_2$.

This is a rather remarkable result in view of the fact that V_∞ is a proper subset of U_2. What this means is that for purposes of ranking pairs of distributions,

[2] See, for example, [3, 4].

the set V_∞ faithfully represents all concave utility functions. To put it somewhat differently, decision makers whose utility functions are in V_∞ can act as spokesmen for the entire group of risk averters; if all decision makers represented by the set V_∞ prefer one distribution to another, then *all* risk averters have the same preference pattern. Thus we see that linearly concave (and single-kinked) functions are an attractive set of functions not only because they simplify the analysis of certain practical decision problems, but also because they constitute a crucial set of functions possessing interesting properties from a purely theoretical point of view.

ACKNOWLEDGEMENT

The author has benefited from several discussions with Tae Kun Seo.

REFERENCES

1. ARROW KJ (1974) The use of unbounded utility functions in expected-utility maximization. *Quart. J. econ*, **88**, 136–138.
2. FLAVELL R (1974) Simple decision problems, a graphical solution. *Omega*, **2**(3), 411–414.
3. HADAR J and RUSSELL WR (1971) Stochastic dominance and diversification. *J. econ. Theory*, **3**, 288–305.
4. HADAR J and RUSSELL WR (1974) Decision making with stochastic dominance: an expository review. *Omega*, **2**(3), 365–377.
5. TILLEY RPR and EILON S (1975) Stochastic dominance for ranking ventures. *Omega*, **3**(2), 177–184.

ADDRESS FOR CORRESPONDENCE: *Professor J Hadar, School of Humanities and Science, Department of Economics, Southern Methodist University, Dallas, Texas 75275, USA.*

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Planning for Teaching–Resource Requirements in Colleges of Technology

N HONEYMAN[1]

JWF POTTS[2]

(*Received January 1975; in revised form October 1975*)

With the increasing demand for higher education, the educational planners need to know more clearly what resources are required to meet the expansion. Teachers are a major cost component and the purpose of this paper is to develop a model to show how various planning decisions affect teaching–resource requirements. The necessity for efficient utilisation of teachers becomes more acute as costs rise rapidly, thus straining the education budget. This study is based on two colleges of technology, one being a large complex college in Dublin, the other being a regional college catering for the much broader regional needs but with a generally lower enrolment in individual courses.

1. FACTORS AFFECTING TEACHING RESOURCE REQUIREMENTS

THERE ARE many factors affecting teaching costs, within the framework of the total costs of a college. The factors are related either to teachers' hours or to course organisation. For simplicity, in this paper, each academic year of a course will be referred to as a course.

Teachers' hours of attendance at a college are comprised of teaching duties and non-teaching duties. Non-teaching duties are comprised of such things as course material preparation, research, liaison with industry, administrative work, etc.

The organisation of courses is affected by such factors as student/staff ratio, class size, student weekly curricular hours, lengths of courses for the academic session and the number of students enrolled for the course. These factors will affect the teaching resources required for a particular course.

Changes in the factors, however, will not always have a linear effect. For instance, small increases in enrolment may merely fill under-utilised classes without requiring formation of a further class group and a consequent increase in teaching–resource requirements. In this case, maximum class size is a relevant

[1] Now with The Royal Bank of Scotland Ltd.
[2] Now with Continental Oil Company Limited.

factor. On the other hand, laboratory, studio or workshop classes may require intensive supervision and increases in enrolment will incur proportional increases in teaching staff. The student/staff ratio applicable to the activity will be the constant of proportionality determining the teaching–resource requirements in this case.

Accommodation and equipment act as constraints on values of parameters. If an acceptable class size is larger than the largest available room, then the class has to be split, thus increasing the number of teachers required. Similarly, possible combination of class groups, from different courses for similar subjects, will be precluded by the non-availability of a large enough room. Availability of specialised equipment or accommodation may restrict class sizes and enrolment.

2. THE MODEL

The output from the model is the full-time equivalent number of teachers of each grade required for a course, a department, a college or for all colleges. The two types of class activity which affect teaching–resource requirements have been mentioned. Firstly, there are those activities for which maximum class size is the relevant factor and which require one teacher per group. Secondly, there are other periods where the number of teaching hours required is proportional to the number of students in the group.

Vol. 4
1976

Consider the first of these activities, i.e. those having a definable maximum class size. A particular course will enrol a certain number of students. These students will comprise one or more class groups. The class group size, and hence the number of class groups may vary during the course of a week. The number of class groups which are required, for a given activity i requiring a grade of teacher j for a course l, will depend on the number of students and the maximum class size as follows:

$$N_{ijl} = \frac{S_l}{C_{ijl}}, \tag{1}$$

where

N_{ijl} is the number of class groups
S_l is the number of students enrolled
C_{ijl} is the maximum class size.

The number of class groups clearly has to be an integer. In practice, a small increase in class size is usually permitted if the formation of a further class group can be avoided. This may be represented by rounding down if the decimal part of the ratio is less than 0·2 and rounding up otherwise.

If there are n maximum class size activities, then the number of hours requiring to be serviced by a teacher of grade j for a course l is given by

$$H_{jl} = D \sum_{l=1}^{n} N_{ijl} SH_{ijl} \tag{2}$$

where

H_{jl} is the annual number of hours needing servicing by a teacher of grade j
SH_{ijl} is student weekly hours for activity i requiring a teacher of grade j
D_l is duration of course in weeks during academic year.

Consider now the second type of activities defined, i.e. those where the number of teaching hours required will rise in proportion to the number of students enrolled. For one course l, the amount of supervision required, for a particular activity i requiring teachers of grade j, will depend on the student/staff ratio for that activity. If there are p activities having different student/staff ratios, then the number of hours requiring to be serviced by a teacher of grade j is given by

$$X_{jl} = D_l \sum_{i=n+1}^{n+p} (S_l/R_{ijl}) PH_{ijl} \tag{3}$$

where

X_{jl} is annual teaching hours of activity requiring servicing
R_{ijl} is student/staff ratio for activity
PH_{ijl} is weekly hours of activity.

Thus for course l, the total number of hours of both types of activity, requiring to be serviced by a teacher of grade j, will be the sum of (2) and (3).

The annual teaching hours for grade of teacher j on course l will then determine the teaching resources required. In practice, teachers are not employed solely on one course, so the teaching resources obtained are in terms of full-time-equivalent teachers and are comprised of part of the teaching time of a number of teachers both full-time and part-time.

Consider a maximum level of teaching hours which is reduced by one or more factors. For one such factor, k, relating to teacher grade j on course l, the reduction would be

$$P_{jkl}\, T_{\max jl}$$

where

$T_{\max jl}$ is maximum level of teaching hours
P_{jkl} is the fraction of maximum teaching hours alleviated.

If there are Z factors reducing maximum teaching hours, then actual teaching hours become

$$T_{jl} = T\max_{jl}\left[1 - \sum_{k=1}^{Z} P_{jkl}\right] \tag{4}$$

The teaching resourses of grade j required for course l are obtained by dividing hours to be serviced by teaching hours, i.e.

$$NT_{jl} = \frac{H_{jl} + X_{jl}}{T_{jl}} \tag{5}$$

where NT_{jl} is teaching–resource reqirement for grade j.

The resource of grade j required, by department, college or for all colleges, is obtained by summing NT_{jl} over the relevant values of l. Teaching costs by grade are obtained by multiplying by average cost per teacher (i.e. salary plus direct overheads).

Vol. 4
1976

3. OPERATION OF MODEL

The model may be used for two main purposes; firstly, to calculate the annual teaching–resource requirements for the range of courses which it is proposed to offer in a college; secondly, for planning purposes by showing the effect of alternative planning decisions on teaching–resource requirements.

The annual teacher requirements for the two colleges in the study were calculated, using the model, for the session 1974–75. The results agreed closely with the traditional manual methods and indicates the usefulness of the model for this purpose.

The model may also be used as a planning tool. An educational development plan is formulated, for example, to increase the number of engineering technicians. All courses of the relevant type will have their enrolments multiplied by a factor which, after allowing for wastage, will give the required increase in students. Initially, only the first years of courses will be affected then other years as students progress. The model will give the increase in teaching–resource requirements from year to year. The planner may then offset the increased requirement for teachers by taking one of a number of courses of action open to him, e.g. reducing student weekly curricular hours, increasing class sizes or student/staff ratios, discontinuing courses, etc. The iterative process between the planner and the model continues until a feasible plan is evolved.

ACKNOWLEDGEMENTS

We thank the Department of Education and the Department of the Public Service for their permission to present this paper and to the many individuals and organisations who provided information pertinent to the study.

ADDRESS FOR CORRESPONDENCE: *JWF Potts, Esq., CONOCO, 105 Wigmore Street, London WIH 0EL, UK.*

1. 4
76

BLANK
PAGE

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Facilities Design with Graph Theory and Strings[1]

JAMES M MOORE

(*Received July* 1975

A facilities design algorithm is developed based upon such graph theory concepts as maximal spanning trees and planar graphs. The upper bound to the demands made by closeness ratings is developed from graph theory. The algorithm lends itself to computer solution minimizing the costs of flow between activities. An alternative approach using strings and the list processing capabilities of computers is suggested. Numerical examples are provided given input in the form of the relationship matrix and working through the block plan layout.

THIS PAPER deals with the problem of facilities design, sometimes called plant layout. The newly emerging branch of mathematics labeled graph theory is utilized. The power of computers has been brought to bear on this problem now for nearly a decade, however, during this time, computer scientists have developed string processors. This paper will attempt to show how string processors can be utilized to assist in facilities design.

The writer assumes that the reader is familiar with the facilities design problem at least to the level described by Muther [4]. Although the concepts from graph theory pertinent to the facilities design problem will be introduced in this paper, the reader with some background in graph theory, perhaps about the level of Harary [2] will find the paper more meaningful. The algorithm described has not yet been implemented on a computer, therefore, the discussion here on string processors is quite shallow so the reader with rather little background can proceed with confidence. The author would welcome comments and observations on this paper from specialists in string processors.

[1] Presented at the combined meeting of the Operations Research Society and the Institute of Management Science on 16 October, 1974 in San Juan, Puerto Rico.

PERTINENT CONCEPTS FROM GRAPH THEORY

Four concepts from graph theory will be utilized in this paper to resolve facilities design problems. They are the concepts of planarity, the dual graph, Euler's formula, and spanning trees. This section is included for the reader unfamiliar with graph theory. The graph theorists can easily skip this section moving on to the example problem where these concepts are utilized.

Planarity

A variety of criteria are known for establishing the condition of planarity. Some theorems characterize planar graphs in terms of forbidden sub-graphs, while other results pin down planarity in terms of the existence of a combinatorial dual graph. Although these criteria are mathematically elegant, the simplest way to understand the condition of planarity is the pictorial approach.

A graph is said to be planar if it can be drawn on a plane surface in such a manner that no two edges intersect. The edges in Fig. 1 shown with solid lines represent a planar graph. If the graph were such that it required the additional edge shown with the dashed line, the graph would be a non-planar graph. The condition of non-planarity suggests that there is no way the graph can be drawn on a surface without at least one pair of edges intersecting. The condition of planarity turns out to be useful in facilities design.

Vol. 4
1976

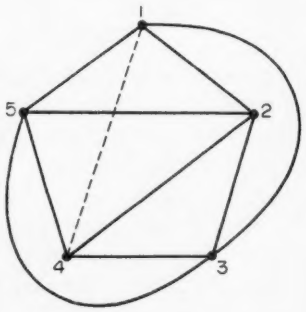


FIG. 1.

Dual graph

One of the special properties of the planar graph is the fact that for each planar graph there exists a corresponding dual graph. Given a planar graph G it is possible to construct its dual graph G^* in the following manner: place a dual vertex in each face of G including the exterior face and, if any two faces have an edge e in common, join the corresponding dual graph vertices by an edge e^* crossing only e. Figure 2 contains a planar graph G shown with solid

lines. The graph shown in dashed lines is the dual graph to G. This relationship between primal graph G and dual graph G^* is useful in facilities design and describing the correspondence of the relationship diagram to the basic floor plans.

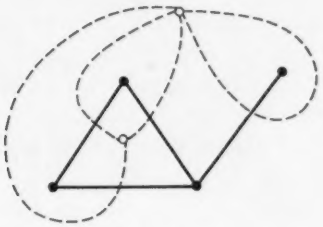


FIG. 2.

Euler's formula

In 1736 topological graph theory was first discovered by Euler and was then dormant for 191 years. Euler's formula was written originally for any spherical polyhedron. It can be restated in terms of the planar graph as follows:

$$V - E + F = 2 \tag{1}$$

where

V = number of vertices,
E = number of edges and
F = number of faces.

Using the planar graph of Fig. 1, that is, omitting the dashed line, verifies Euler's formula when we observe $5 - 9 + 6 = 2$. The reader is left to verify for himself that both the primal graph with solid lines and the dual graph with dashed lines in Fig. 2 follow Euler's formula.

Figure 1 happens to depict a condition of particular interest in facilities design. Figure 1 represents not only a planar graph but a special subset of planar graphs namely, it is a maximal planar graph. The maximal planar graph is one to which no line can be added without losing planarity.

It can be shown that if a graph is a maximal planar graph, then every face is triangular. This is to say that each face, including the exterior face, has exactly three edges. Furthermore, Seppanen [5] has shown that the maximum number of edges in a planar graph is

$$E_{max} = 3V - 6. \tag{2}$$

This relationship is important in facilities design because it defines the upper bound on the number of closeness relationships that can be physically satisfied in a two dimensional layout.

Spanning trees

In graph theory terminology, a tree is a connected graph which contains no closed loops. The solid lines in Fig. 3 provide an illustration of a tree. The primal graph with solid lines in Fig. 2 is not a tree because it contains a closed loop. Of special interest in facilities design is a particular type of tree called the spanning tree. The spanning tree of graph G, shown in solid lines, is a subgraph in the form of a tree which contains all of the vertices of G. A graph may have a large number of spanning trees. Particularly useful in facilities design is the maximum spanning tree. When graph G has weighted edges, the maximum spanning tree is that spanning tree the sum of whose edges is a maximum. Having defined these graph theory concepts, let us move on to an example problem.

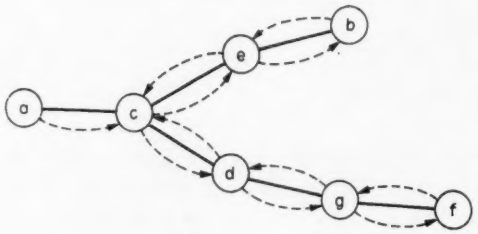


FIG. 3.

ILLUSTRATIVE PROBLEM

For ease of discussion, let us consider a small perhaps almost trivial problem from the layout planner's point of view. Seven activities ($n = 7$) from the office at Brown Scientific need to be arranged in a layout. Let us label these activities a through g. With seven activities we have 21 pairs of closeness priorities [$n(n - 1)/2$] which have been determined by management.

Let us use Muther's [4] rather widely accepted categories for closeness priority, namely where closeness is (1) absolutely necessary, (2) extremely important, (3) important, (4) ordinary closeness ok, (5) unimportant, and (6) undesirable. Management has placed a closeness priority on the 21 pairs of relationships as follows:

A	Absolutely necessary:	*ac.*
E	Extremely important:	*ce, dg, fg.*
I	Important:	*ae, be, cd.*
O	Ordinary closeness:	*ab, ad, af, ag, bc, bf, bg, cf, cg.*
U	Unimportant:	*de, df, ef, eg.*
X	Undesirable:	*bd.*

Vol. 4
1976

Find maximal spanning tree

Given this data, the first step is to determine the maximal spanning tree. We will use Kruskal's algorithm [3] because it is simple to follow manually. It states:

> Among the edges which are not yet part of the tree, choose the edge which has the largest value and which does not form a closed loop with those already selected.

For implementation on the computer, there are a variety of other algorithms for locating the maximal spanning tree. Whitney [6] has recently published a FORTRAN program to locate the minimal (or maximal) spanning tree which is based upon a technique suggested by Dijkstra [1].

The Kruskal algorithm guarantees an optimal solution. It is possible there may be more than one optimal solution but optimality is guaranteed. Table 1 depicts the developing tree in the form of a string. All but the last string contain one activity which is underlined. It can be observed that each underlined activity will appear twice one space apart in the rewriting of the next string as the tree develops. That is, the underlined activity will be rewritten in the string contained in the following step. Kruskal's algorithm terminates when all activities have been included in the tree which takes 6 steps in the case of our problem, that is, one step less than the number of activities in the problem.

Although strings are easy for the computer, they are somewhat difficult to picture. Fig. 3 depicts the maximal spanning tree developed by Kruskal's algorithm. It is equivalent to the string developed in Step 6 in Table 1.

TABLE 1. MAXIMAL SPANNING TREE DEVELOPMENT UTILIZING STRINGS

Step	Edge being added	Developing tree	Comments
1	*ac*	*a<u>c</u>*	
2	*ce*	*a<u>c</u>ec*	
3	*be*	*a<u>c</u>ebec*	Note that neither *dg* nor *fg* can be added at this time because they are not connected to the trunk at this point. Furthermore *ae* cannot be added because it would create a closed loop.
4	*cd*	*ac<u>d</u>cebec*	
5	*dg*	*acd<u>g</u>dcebec*	
6	*fg*	*acdgfgdcebec*	Maximal spanning tree reached.

Edge adding algorithm

The edge adding algorithm is really just a heuristic. It does not guarantee an optimal solution, however, it does provide a mechanism whereby we may systematically be able to add edges in such a manner as to triangulate the tree.

A graph is said to be triangulated if each of its faces has three sides (edges). If this can be achieved, our solution is at its upper bound as defined by Euler. There is no point in working further.

Basically, the heuristic first introduces the edges with highest closeness priority. In the case of ties where more than one edge at the same priority level are still outside the solution, we introduce the edge whose activities are closest in the existing string. In the string we will depict triangulation with three letter words. Since our goal is to triangulate the whole string, the one exception to the tie breaking rule is when the edge has already been added to solution as part of a three sided face. After each edge is added, the graph must be tested for planarity. The edge adding algorithm ceases when (1) either all high priority edges have been added, or (2) a maximal planar graph is developed, or (3) the graph is triangulated, i.e. all faces have exactly three edges. Table 2 takes the maximal spanning tree and adds edges to it using this heuristic. Table 2 ends with the addition of edge *ag* when all the words in the string have become three letter words. Graph *G* is triangulated, i.e. all its faces have three sides.

Vol. 4
1976

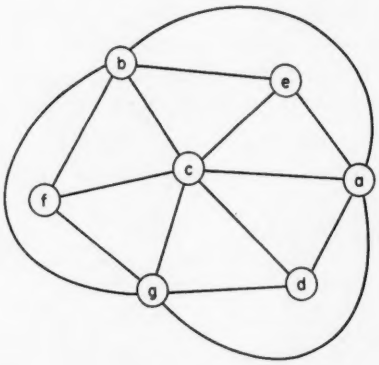


FIG. 4.

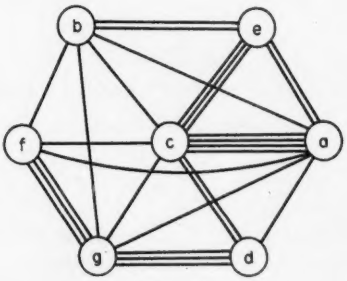


FIG. 5.

In Table 2 each string has a pair of activities underlined. The underlined activities represent the next edge to be added into the string. The final string suggests triangulation has taken place. Figure 4 is the result of taking the final string from Table 2 and converting it to pictorial format. It can be seen that each of the faces in Fig. 4 have three sides, that is to say they are triangulated. Figure 5 depicts a relationship diagram prepared using the traditional method of systematic layout planning (SLP). Using Muther's [4] convention, the four lines suggest an *A* rating, three lines an *E* rating, two lines an *I* rating, and one line an *O* rating. Similarity between Fig. 4 and Fig. 5 which was prepared using SLP is quite striking.

TABLE 2. SUMMARY OF EDGE ADDITION HEURISTIC

Adding this edge	Because	Changes the string to read
		acdgfgdcebec or *cdgfgdcebeca*
ae	Highest priority (important) not yet entered in solution.	*cdgfgdcebea eca*
cg	From ordinary ratings has min activities in the interval, i.e. 1.	*cdg cgfgdcebea eca*
cf	Min activities, 1, in interval.	*cdg cgf cfgdcebea eca*
bc	Min activities, 1, in interval.	*cdg cgf cfgdcb cebea eca*
ab	Min activities, 1, in interval.	*cdg cgf cfgdcb ceba bea eca*
bg	For min interval *ad*, each activity already triangulated, so introduce *bg* having next smallest interval, i.e. 2.	*cdg cgf cfgb gdcb ceba bea eca*
bf	Minimum interval	*cdg cgf cfb fgb gdcb ceba bea eca*
ad	Skipping those already triangulated, *ad* has min interval of 5.	*cdg cgf cfb fgb gda dcb ceba bea eca*
ag	Among those not yet triangulated, min interval is 8.	*cdg cgf cfb fgb gda dcb ceb aga bea eca*

Stop, because all words in the string are three letter words, i.e. the string is triangulated.

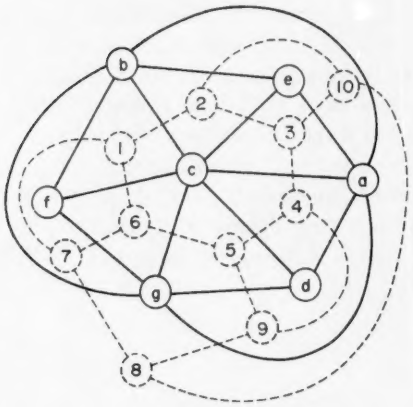


FIG. 6.

Dual graph preparation

In the facilities design problem, one attempts to depict the arrangement of the activities by defining the borders of the activities involved. In the illustrative problem, activities *a* through *g* need therefore to be delimited. The dual graph has the property of delimiting each of the nodes of the corresponding primal graph. Using the rule for constructing a dual graph described earlier, Fig. 6 depicts the dual graph G^* which corresponds to the primal graph G depicted in Fig. 4. Care must be taken to make sure the dual graph has a vertex in the infinite face of the primal graph. The dual graph delimits each of the activities in the primal and therefore can be viewed as a preliminary layout plan.

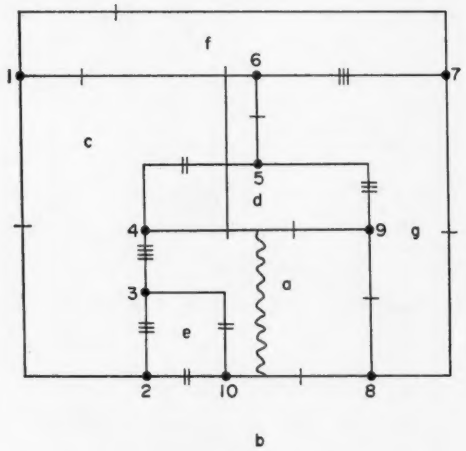


FIG. 7.

Assuming that the area requirements are given for each of the activities, one can now take the dual graph of Fig. 6 and combine it with area requirements to develop a block plan layout. Figure 7 depicts such a block plan layout developed from the dual graph. Notice that wherever three departments come together we find one of the dual graph vertices. Figure 7 represents only one of a number of possible layout plans depending on the area requirements for each of the activities. From Fig. 7 we see that it is possible to generate a block plan layout utilizing some graph theory concepts and some concepts of strings.

EVALUATION

Figure 7 contains a crude graphical evaluation of the effectiveness of the methodology. The four lines crossing the common border between *a* and *c* indicate an *A* or absolutely necessary closeness rating. The closeness rating lines crossing between activities use the same coding as contained in Fig. 5. We observe that all but one high priority rating has a common border. The exception is the long distance between activity *a* and *f* which was the edge omitted by the heuristic when it terminated. The wavy line between *d* and *b* suggests that the undesirable relationship between *b* and *d* are spread apart by some distance.

This methodology appears to be capable of generating an acceptable solution. Perhaps the most notable drawback to the methodology is the graph theory demand of including the infinite face. This demand causes one of the activities to be the exterior face in the block plan layout, activity *b* in our illustrative example. A suggestion to avoid this limitation would be to add a dummy activity into the original data with very low priority ratings. One would speculate that if its priority ratings were very low, the methodology might force it into the infinite face.

Perhaps the methodology can be most easily summarized in the format of a flow chart. Figure 8 depicts an overview flow chart. Although this heuristic has not been implemented on a computer, computer people with experience with list processors are invited to consider this problem. It would appear that being able to describe the facilities design problem in the form of strings would lend itself to being solved by a list processor. Comments from specialists in list processing would be welcomed by the author.

SUMMARY

This paper has extracted some basic concepts from the mathematics of graph theory with the intent of bringing them to bear on the facilities design problem. The graph theory concepts involved are planarity, the dual graph, Euler's

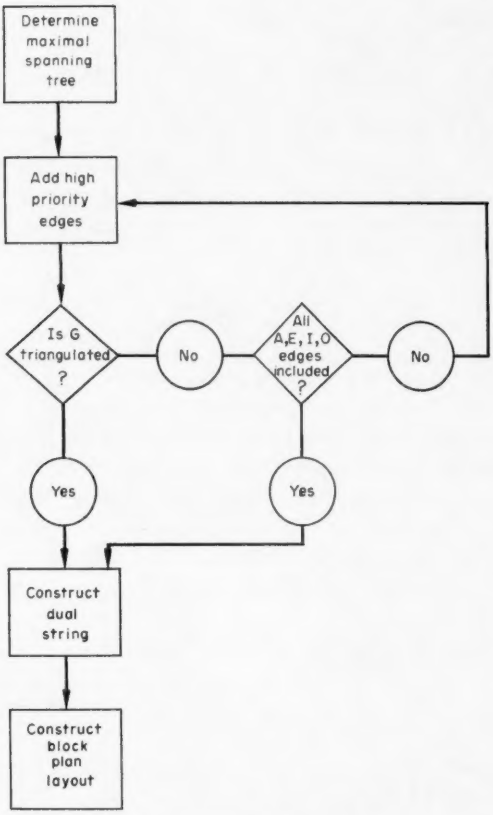


FIG. 8. *Overview of the approach.*

formula, and spanning trees. The methodology proposed is illustrated with a small facilities design problem. Given a set of data defining closeness priorities, the methodology first determines the maximal spanning tree after which a heuristic is introduced to add edges until either a maximal planar graph is developed or all of the high priority edges are included. The use of strings to define the various stages of the graph is recommended. The dual graph is prepared which can be converted into a block plan layout. The author recommends further research testing the feasibility of the methodology for computer solution.

REFERENCES

1. DIJKSTRA EW (1959) A Note on two problems in connection with graphs. *Numerical Mathematics*, 5(1), 269–271.

2. HARARY FRANK (1969) *Graph Theory*. Addison-Wesley, Reading, MA, U.S.A.
3. KRUSKAL JB, JR (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* 48–50.
4. MUTHER RICHARD (1973) *Systematic Layout Planning* (2nd Edition). Cahners, Boston, MA, U.S.A.
5. SEPPANEN J and MOORE JM (1970) Facilities planning with graph theory. *Mgmt Sci.*, **17**(4), B–242.
6. WHITNEY V and KEVIN M (1972) Algorithm 422 minimal spanning tree. *Communications of the ACM* **15**(5), 273.

ADDRESS FOR CORRESPONDENCE: *Professor JM Moore, Department of Industrial Engineering and Operations Research, College of Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

An Efficient Heuristic Procedure for Solving the Layout Design Problem

GG HITCHINGS

M COTTAM

(*Received June 1975; in revised form November 1975*)

The limited ability of schematic procedures, constraints of linear programming techniques, inflexibility of construction methods and inadequacy of dynamic programming approaches to provide acceptable solutions to realistic size layout design problems has led to an ever increasing interest in heuristic procedures. Comparative studies have shown that heuristic procedures can satisfy more desirable qualities in an 'ideal algorithm' to a greater extent than competitive techniques. Excessive computational effort, which has been one of the main criticisms levelled against heuristic approaches in the past, proves to be less of a problem in relation to layout design. By utilizing unique attributes associated with real-life problems and using small incisive modifications it is demonstrated how a heuristic procedure can be significantly improved.

Vol. 4
1976

INTRODUCTION

A HEURISTIC technique may be defined as a method ". . . . for solving problems by an intuitive approach in which the structure of the problem can be interpreted and exploited intelligently to obtain a reasonable solution . . . ." Nicholson [9]. Heuristic procedures lend themselves ideally to fast computing methods, usually involving the scanning of many alternative solution attempts and, by defined criteria, selecting the better or best of these solutions. Nicholson indicates also that '. . . . heuristic techniques are the only way of tackling many large optimization problems . . . .' Practical layout design problems fall into this category.

Much previous work on the layout design problem has been brought together under the heading 'Location—Allocation Systems—An Annotated Bibliography', AC Lea [8], where something in excess of one thousand references have been summarized covering a period of approximately forty years. A large proportion of these references has appeared only in the last ten years, approximately 86 per cent. Interest has been considerable in heuristic procedures and about 450 references have been identified in which use has been made of such approaches.

The reason for continuing interest in heuristic algorithms for solving the ayout design problem has been hinted at by Nicholson with reference to problem size, see paragraph 1 above. El-Rayah and Hollier [3] in their review of Plant Design Techniques, highlight some of the inadequacies of alternative approaches for solving the layout design problem, viz. schematic approaches, when too subjective a method is entailed, resulting in considerably less than optimal solutions. Graphical and systematic procedures give rise to similar problems as schematic techniques, but at least attempt to optimize within given constraints. Linear programming methods will provide optimal solutions to the facilities design problem for imposed constraints but cannot be regarded as a general method of solution. Treated as a quadratic assignment problem, optimal solutions are only possible for the smallest of problems, normally n, number of facilities to be assigned to locations being < 12. As stated, these various limitations on solution method have led to an increasing interest in heuristic procedures.

Heuristic procedures

Probably the most important of all heuristic procedures devised for solving the layout design problem is CRAFT, Computerized Relative Allocation of Facilities Technique, originated by Armour and Buffa [1]. The basic heuristic of CRAFT is that of evaluating from a given assignment of facilities to locations, improved assignments which, on an iterative basis, improve the flow characteristics of the layout design. Criterion for improvement of flow characteristics is progressive reduction of materials handling cost in terms of distance and quantities moved.

Improvement of the layout design is brought about by consideration of $n(n-1)/2$ possible pairwise—exchanges of facilities to locations, acceptance of superior and rejection of inferior assignments. The process is repeated until no further improvement in flow characteristics is possible.

Many subsequent computing procedures have been based on this original pairwise-exchange heuristic rule. They include:

(i) H 63 by Hillier [4],

(ii) HC 63–66 by Hillier and Connors [5],

(iii) COL—Computerized Office Layout by Vollman, Nugent and Zartler [11],

(iv) Biased Sampling by Nugent, Vollman and Ruml [10],

(v) MAT—Modular Allocation Technique by Edwards, Gillet and Hale [2],

(vi) FRAT—Facilities Relative Allocation Technique by Khalil [7].

The six mentioned were chosen for comparison purposes as they all stem from the basic heuristic used in the CRAFT procedure. A technique devised by the authors [6], 'Terminal Sampling Procedure' is used to show how incisive modification of the heuristic procedure can improve considerably on the basic algorithm. In essence, TSP is an amalgam of previous heuristic algorithms embracing many of the desirable elements which would tend to contribute towards an ideal algorithm. Attributes of TSP which make it probably one of the more advanced of techniques available at present are:

(a) its ability to make better use of expended computing time and,
(b) its ability to produce better solutions to practical layout design problems.

It would be useful at this stage to summarize the usual criticisms levelled at heuristic procedures. They are:

(1) Normally extremely time consuming in computing effort.
(2) Will not guarantee optimal solutions to problems.
(3) Solution methods are unpalatable to the pure mathematician.

Vol. 4
1976

Terminal sampling procedure (TSP)

TSP sets out to minimize time consumed in computing effort, produce as good, or better solutions than the 'best' previous techniques. A singular consolation concerning mathematic purity of solution is that one is dealing with a real-life problem which in fact is harder to define than actually solve! Problems solved are those by Nugent, Vollman and Ruml [10]. These problems are now classical in the literature dealing with solution methods for the layout design problem. They constitute a set of eight problems ranging from five facilities to be assigned to five locations up to thirty facilities to be assigned to thirty locations. Since 1968 these problems have presented a challenge to layout design specialists and provide a comparative basis for both quality and economy of computing time in solution methods.

Small incisive changes in the heuristic procedure which make TSP more effective are:

(a) Pairwise-exchange is carried out on a selective basis by only considering those distances and quantities flowing which contribute mostly to the assignment value, i.e. the highest product vector embracing two facilities/locations. It is these two which are involved in the initial pairwise-exchange routine—compare COL [11]. Primarily computing effort is reduced from $n(n-1)/2$ exchanges down to $2n-4$ exchanges.
To ensure that maximum improvement in solution quality is achieved each iterative $(2n-4)$ path is terminated by a CRAFT loop.

Using input data from the largest of the classical eight problems, i.e. the 30 × 30 problem, Fig. 1 shows how CRAFT and TSP compare in ability to reduce assignment cost value with respect to computing time consumed. Note for example that merely incorporating this minor modification, TSP reduces an initial input value of 4030 to 3300 in 135 computer processor unit seconds whereas the basic CRAFT routine takes 325 cpu seconds to reduce the value by the same amount. In practical terms this means that for larger size problems using a cut-out after a timed run, TSP will give significantly better solutions than CRAFT.

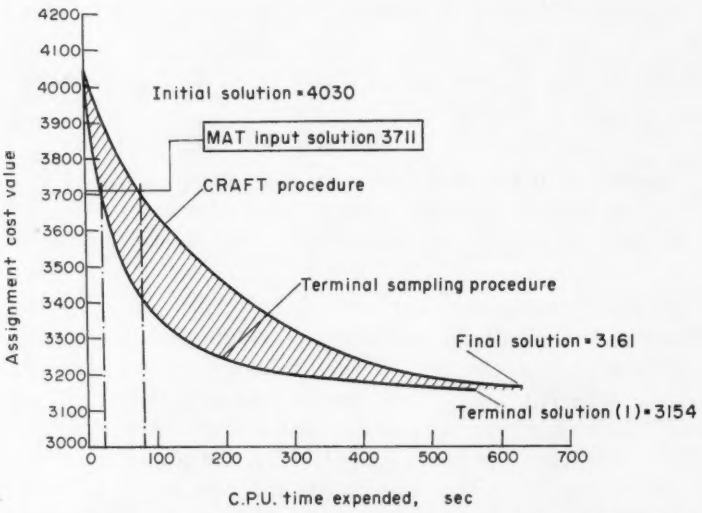


FIG. 1. *Graph showing comparison between CRAFT and terminal sampling procedure on a timed computer run.*

(b) TSP has the ability to use improved input solutions—compare MAT [2]. In the case of TSP however, input solutions are not generated separately as for MAT but obtained by identifying 'tie' assignment values down the iterative path.

Figure 2 shows how these tie values are identified and indicates the subsidiary paths followed down the 'tree'. Note that the new input solution of 3386 produces a final solution of 3086. Graph Fig. 3 shows how other input solutions constitute the 'sample' in TSP. Comparing Fig. 1 and Fig. 3 it may be seen that the saving effected by inputing a solution of MAT value 3711 is insignificant. Only significant savings in cpu time can be gained by inputing solutions of much better value, as is possible with TSP.

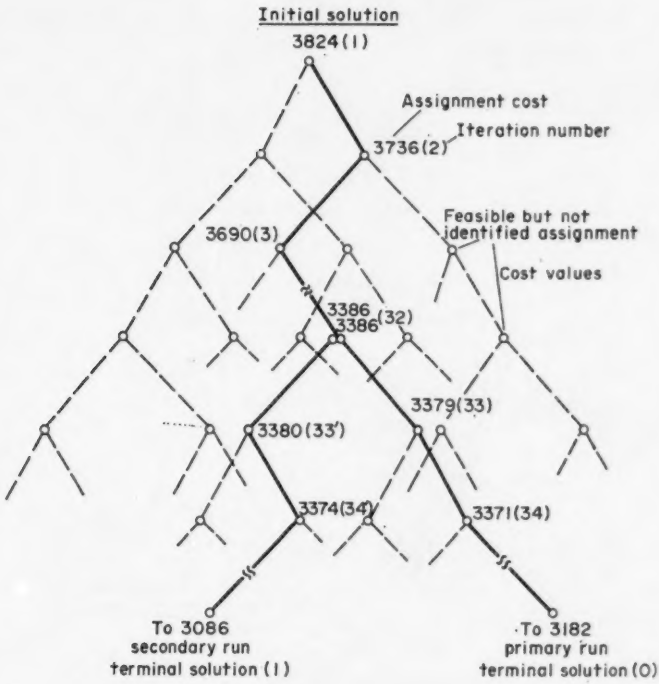


FIG. 2. *Terminal sampling procedure showing tree-searching method for establishing secondary solutions by sampling tie values generated from a primary run. An initial input solution of* 3824 *reduces to* 3182 *on a primary run whilst a secondary run at tie value* 3386 *iterates down to* 3086.

(c) In common with other heuristic methods, TSP retains the ability to 'fix' facilities in locations. In practice this allows a fairly large size problem to be handled by effectively reducing the size of it, i.e. many facilities need to be fixed in position for example, to be adjacent to fume extraction equipment, easy access to 'goods outward' bays, etc.

Analysis of real-life problems indicates that the Pareto Principle operates. When a layout design problem can be represented by matrices detailing distance and flow requirements, it is noticeable that a significantly small number of cells in the Quantity-Flow matrix describe the majority of total flow in the system. See Fig. 4 which is the result of a study carried out in a company producing machined die-castings. This curve is typical of many companies involved in batch and mass production industries.

Vol. 4
1976

What this means in computing time saved can best be illustrated using the 30 × 30 problem again.

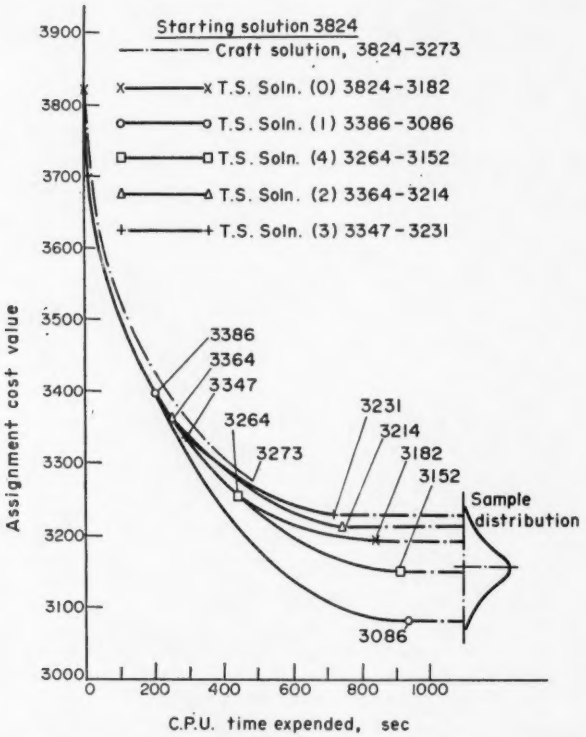


FIG. 3. *Graph showing terminal sampling procedure. From a starting solution of* 3824, *run to a final solution of* 3182, *samples are taken at tie values ard run to terminate at solutions shown.*

Figure 5 shows that if only five of the thirty facilities were fixed in position then a saving in computing effort can amount to 50 per cent. It is not unusual in practical problems to be able to 'fix' a higher percentage than this.

Tables 1 and 2 show comparisons made between the various heuristic procedures and TSP. Table 1 shows that TSP produces solutions of quality equal to or better than the Biased Sampling Technique [10], which hitherto has acclaim of being the technique which produces highest quality solutions. Table 2 shows that computing time need not be prohibitive, as would seem to be the case with Biased Sampling. In a timed computer run it was demonstrated that by judicious 'fixing' of twenty facilities in the 30 × 30 problem a value of 3818 was reduced to 3078 in just 37 cpu seconds. A considerable improvement on Biased Sampling. See Tables 1 and 2.

Vol. 197

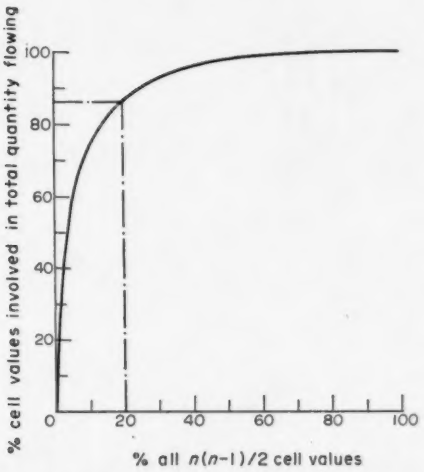


FIG. 4. *Graph showing Pareto Principle in operation for total flow involved in an actual factory layout. Note that some* 20% *of* n(n−1)/2 *cells in the flow contain values representing* 86% *of total flow.*

Vol. 4
1976

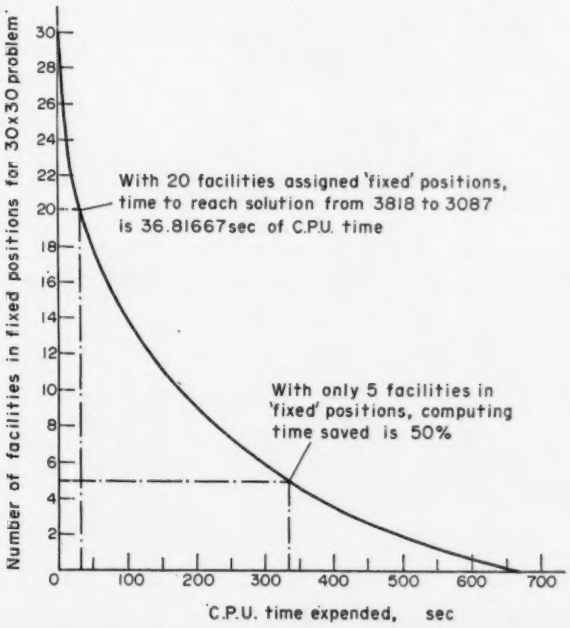


FIG. 5. *Graph showing reduction in CPU time expended as a result of fixed facilities in positions for* 30 × 30 *problem. From all* 30 *facilities in fixed positions to all facilities floating.*

TABLE 1. TABULATED RESULTS OF A COMPARISON BETWEEN TERMINAL SAMPLING PROCEDURE AND OTHER HEURISTIC ALGORITHMS FOR ABILITY TO PRODUCE SOLUTIONS OF HIGH QUALITY. 8 PROBLEMS BY NUGENT *et al.* (1968).

No. of facilities *n*	Random sampling	H 63 (Hillier) 1963	HC 63–66 (Hiller & Connors) 1966	MAT (Modular allocation technique)	CRAFT (Computerized relative allocation of facilities technique)	COL (Computerized office layout)	FRAT (Facilities relative allocation technique)	Biased sampling technique	Terminal sampling procedure	Optimal solution (where known)
5	29	25	29	25	25	25	25	25	25	25
6	43	43	43	55	43	43	43	43	43	43
7	82	78	74	80	74	74	74	74	74	75
8	116	109	107	128	107	107	107	107	107	107
12	353	301	304	337	289	293	295	289	289	?
15	700	617	578	741	583	583	575	575	575	?
20	1505	1384	1319	1450	1324	1312	1300	1304	1296	?
30	3847	3244	3161	3711	3148	3152	3129	3093	3086	?

TABLE 2. TABULATED RESULTS OF COMPUTING TIME EXPENDED IN SECONDS PER SOLUTION BY DIFFERENT ALGORITHMS. 8 PROBLEMS $n = 5$ to $n = 30$ NUGENT *et al.* (1968).

No. of facilities n	Random* sampling	H 63* (Hillier) 1963	HC 63–66* (Hillier & Connors) 1966	MAT* (Modular allocation technique)	CRAFT* (Computerized relative allocation of facilities technique	COL† (Computerized office layout)	FRAT‡ Facilities relative allocation of facilities technique	Biased* sampling	Terminal§ sampling procedure
5	0.12	6.0	10.0	1.0	1.0	3	0.11	1.1	0.25
6	0.16	7.0	9.0	1.3	2.0	4	0.15	2.1	0.30
7	0.22	15.0	12.0	1.6	5.0	5	0.25	5.7	0.75
8	0.28	14.0	14.0	2.1	10.0	6	0.33	10.9	1.23
12	0.65	55.0	19.0	6.1	70.0	8	1.25	65.8	6.84
15	0.95	78.0	40.0	13.2	160.0	19	2.96	219.2	20.13
20	1.67	168.0	75.0	47.5	528.0	32	10.87	691.5	59.87
30	4.09	398.0	285.0	346.7	3150.0	118	58.30	4272.4	383.27

* Ran on G.E. 265 computer (MAT is conversion time from IBM 7094 machine factor ÷ 5).
† Ran on G.E. 635 computer.
‡ Ran on IBM 360 computer.
§ Ran on I.C.L. 470 computer.

IN CONCLUSION

Heuristic procedures provide a practical method of analysing and solving many problems met in manufacturing industry. Apart from layout design problems, other areas of use include line balancing, site location, scheduling, routing, design problems and many others. Criticism has been heavy, usually because of excessive computation time entailed in applying the heuristic procedure. It has been domonstrated that computational effort need not be a problem if the heuristic procedure is examined in some depth and judicious modification carried out. One would hope that heuristic methods which have been suggested in the past are not easily discarded on the basis that they are too time consuming to use, but re-examined, suitably modified in the light of present findings and take their rightful place as useful management tools.

The authors of this paper are interested in high quality solution methods to the layout design problem whether by heuristic or other means. They are particularly interested in solutions and assignment plans for those problems by Nugent *et al.* [10]. Any researcher who has produced better solutions than those quoted in this paper please contact the authors.

REFERENCES

1. Armour GC and Buffa ES (1963) A heuristic algorithm and simulation approach to relative location of facilities. *Mgmt Sci.* **9**(2), 294–309.
2. Edwards HK, Gillet WE and Hale ME (1970) Modular allocation technique. *Mgmt Sci.* **17**(3), 161–169.
3. El-Rayah TE and Hollier RH (1970) A Review of plant design techniques. *Int. J. Prod. Res.* **8**(3), 263–279.
4. Hillier FS (1963) Quantitative tools for plant layout analysis. *J. Ind. Engng* **14**(1), 33–40.
5. Hillier FS and Connors MM (1966) Quadratic assignment problem algorithms and the location of indivisible facilities. *Mgmt Sci.* **13**(1), 42–57.
6. Hitchings GG (1973) Analysis and Development of Techniques for Improving the Layout of Plant and Equipment. Doctoral Thesis—University Wales, Cardiff, U.K. *Dept. Mech. Engng and Engng Prodn.*
7. Khalil TM (1973) Facilities relative allocation technique. *Int. J. Prod. Res.* **2**(2), 183–194.
8. Lea AC (1973) Location—Allocation Systems—An Annotated Bibliography. University of Toronto, Dept. of Geography. Discussion Series, Paper No. 13.
9. Nicholson TAJ (1971) *Optimization in Industry Vol.* 1. *Optimization Techniques.* Chap. 10, pp. 152–160. Longman Press, London. Business School Series.
10. Nugent CE, Vollman TE and Ruml J (1968) An experimental comparison of techniques for the assignment of facilities to locations. *J. Ops. Res. Soc. Am.* **16**(1), 150–173.
11. Vollman TE, Nugent CE and Zartler RL (1968) A computerized model for office layout. *J. Ind. Engng* **19**(7), 321–327.

ADDRESS FOR CORRESPONDENCE *Dr. Graham G. Hitchings, Department of Mechanical Engineering and Engineering Production, University of Wales Institute of Science and Technology, Cathays Park, Cardiff, UK.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Management Applications of the Inverse Gaussian Distribution

GA WHITMORE

(*Received September 1975; in revised form November 1975*

The inverse gaussian distribution is a relatively unknown statistical distribution which appears to have considerable potential for application in the management sciences. The paper presents a summary of the distribution's key statistical properties and surveys applications of the distribution found in the literature. The paper also discusses a number of areas of potential application which are as yet unexplored.

Vol. 4
1976

THE INVERSE gaussian distribution is receiving increasing attention in both the applied and theoretical statistical literature as evidenced by the extensive bibliography of this paper. In management and the other social sciences, it has received less attention possibly because it is not widely known. In this paper, some of the more important properties of this distribution family are presented and a number of areas are described where it has been or, perhaps, can be applied. Of necessity, many details must be omitted from this paper. For the reader who wishes to investigate specific aspects of the distribution or its applications in greater depth, he is referred to the relevant literature sources cited in the paper.

THEORETICAL PROPERTIES AND CHARACTERISTICS

It will be seen later that the *inverse gaussian distribution* is frequently associated with duration phenomena, so it is suggestive to denote its random variable by T. The probability density function of T is:

$$f(T) = (\lambda/2\pi T^3)^{\frac{1}{2}} \exp[-\lambda(T-\mu)^2/2\mu^2 T] \qquad (1)$$

where $T > 0$ and λ and μ are positive parameters. By setting $\mu = 1$ in (1), a distribution often referred to as the *Wald distribution* is obtained.[1]

[1] The reference to Wald is due to his derivation of this type of distribution as the limiting form of the distribution of sample size in certain sequential probability ratio tests [33].

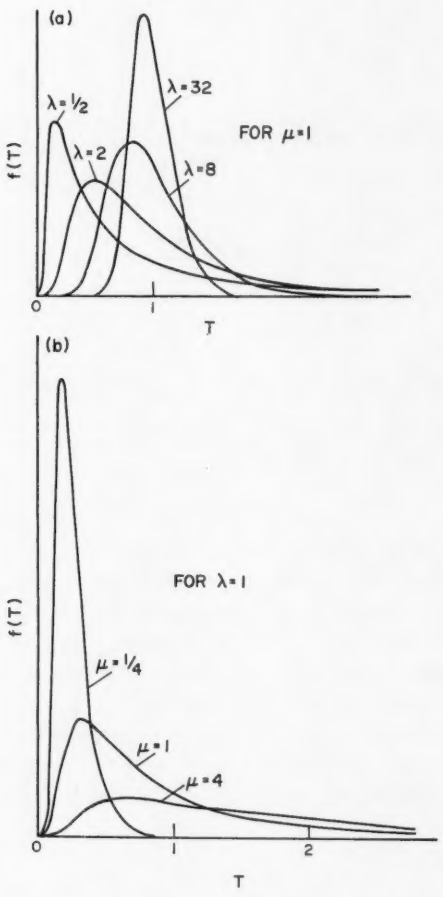


FIG. 1.

The mean and variance of the inverse gaussian distribution are μ and μ^3/λ, respectively. It is a unimodal, right-skewed distribution with a shape which varies widely as μ and λ vary (see Fig. 1). Its mode is located at:

$$\mu\left\{\left[1+\left(\frac{3\mu}{2\lambda}\right)^2\right]^{\frac{1}{2}}-\frac{3\mu}{2\lambda}\right\}. \tag{2}$$

If a simple random sample of size n is taken from an inverse gaussian distribution, and M and H denote the sample arithmetic mean and sample harmonic

mean, respectively, then the joint maximum likelihood estimators of μ and λ are M and L, respectively, where:[2]

$$\frac{1}{L} = \frac{1}{H} - \frac{1}{M}. \tag{3}$$

The estimators, M and L, are jointly sufficient and independent estimators of μ and λ. The sampling distribution of M is itself inverse gaussian with $\mu_M = \mu$ and $\lambda_M = n\lambda$. The sampling distribution of L is of the chi-square type; specifically $n\lambda/L$ has a chi-square distribution with $n - 1$ degrees of freedom.

Probability values for the inverse gaussian distribution may be computed from:

$$1 - F(T) = N[-(T-\mu)\sqrt{(\lambda/\mu^2 T)}] - \exp[2\lambda/\mu]N[-(T+\mu)\sqrt{(\lambda/\mu^2 T)}] \tag{4}$$

where $F(T)$ is the distribution function of T and $N[\cdot]$ is the standard normal distribution function. Letting $\theta = \lambda/\mu$ and employing the standardizing transformation $W = T/\mu$, (4) may be rewritten as follows:

$$1 - G(W) = N[-(W-1)\sqrt{(\theta/W)}] - \exp[2\theta]N[-(W+1)\sqrt{(\theta/W)}] \tag{5}$$

where $G(T/\mu) = F(T)$.

To the author's knowledge, accessible published tables for the inverse gaussian distribution are not available. It is reported in [8, p. 151] that Krapivin [10] has published (in Russian) tables of the Wald distribution. Wasan and Roy reported in an abstract [34] that they have tables for the cumulative inverse gaussian distribution. In [41], Yalovsky and the author have computed values of $G(W)$ to four decimal places for W values ranging over 0(0·01)4(0·1)10(1)39 and θ values ranging over 0·05(0·05)10·05. [41] also contains a table of selected percentiles for $G(W)$. Probabilities for the limiting distribution obtained as $\mu \to \infty$ (and, hence, $\theta \to 0$) are computed readily from standard normal tables because (4) reduces to the following simple form in that special case:[3]

$$1 - F(T) = N[\sqrt{\lambda/T}] - N[-\sqrt{\lambda/T}]. \tag{6}$$

For large values of θ, the inverse gaussian distribution may be approximated by the normal distribution, although θ must be very large for the approximation to be close in the tail areas.

For more complete theoretical documentation on the inverse gaussian distribution than is given here, the reader is referred to [31, 32, 36, 8].

[2] As a practical matter, the reader should note that M will be sensitive to measurement errors in large values of T and H will be sensitive to measurement errors in small values of T. Furthermore, where sample data are grouped into classes, the maximum likelihood estimates will be sensitive to assumptions made about observations in the first and last classes in the sample frequency distribution.

[3] This limiting distribution is used in [16], for instance.

Vol. 4
1976

RELATION TO A WIENER (BROWNIAN) DIFFUSION PROCESS

A statistical characterization of the inverse gaussian distribution which is useful for identifying potential applications of the distribution and for understanding its properties is provided by a Wiener (or Brownian) diffusion process (cf. [2, pp. 203–251]).

Consider a stochastic process $H(t)$ which has level a at time t_A. As time passes, the level of the process fluctuates randomly with a tendency to drift towards an absorbing barrier located at level b ($b > a$). At some random time t_B, the process reaches level b for the first time. This first passage time T to the absorbing barrier (i.e. $T = t_B - t_A$) is the random variable of interest. If the stochastic process is a Wiener diffusion process with drift parameter $\delta > 0$ and volatility parameter $\nu > 0$, then T has an inverse gaussian distribution with $\mu = (b - a)/\delta$ and $\lambda = (b - a)^2/\nu$. The conditions which must be satisfied by $H(t)$ for it to be a Wiener process are the following:

I. For non-overlapping time intervals (t_1, t_2) and (t_3, t_4), the random variables $H(t_1) - H(t_2)$ and $H(t_3) - H(t_4)$ must be statistically independent.

II. For any time interval (t_1, t_2) with $t_2 > t_1$, $H(t_2) - H(t_1)$ must be normally distributed with mean $(t_2 - t_1)\,\delta$ and variance $(t_2 - t_1)\nu$.

Vol.
197

APPLICATIONS

To date, management and other social science applications of the inverse gaussian distribution have been restricted primarily to studies of duration phenomena. Nevertheless, since duration phenomena are so pervasive in the social sciences, the potential importance of the distribution is by no means insignificant. The paper now turns to summaries of several of these applications. Of necessity, the discussion of each is brief—the intent simply being to make the reader aware of areas of potential application.

Duration of a strike (Lancaster [11])

In this paper, the author studies the durations of strikes in the United Kingdom based on data recorded by the Ministry of Labour commencing in 1965. The abstract of [11] describes the conceptual model employed, i.e. "The settlement of a strike is regarded as a probabilistic process and the duration of a strike is treated as an observation on a random variable. A model for this random duration is created by supposing that at each point of time after the commencement of a strike there exists an index of the difference between the parties to the dispute. This index is itself regarded as a one-dimensional stochastic

process in continuous time and space and determines the duration of the strike by the first point of time at which the difference index passes through an absorbing barrier representing 'agreement'. The duration of a strike then becomes the First Passage time of a stochastic process to a single absorbing barrier. As a first approximation the index process is assumed to be simple Brownian motion with drift and in consequence the duration of a strike has the Inverse Gaussian for its probability distribution".

The author proceeds to fit the inverse gaussian distribution to observed strike-duration data for several industry groups, obtaining as a byproduct, estimates of the drift and volatility parameters of the associated Brownian motion. In each case, he tests the goodness of fit by means of a chi-square statistic and, as he reports, the fit of the inverse gaussian distribution to the observations is, with few exceptions, very close.

Length of hospital stay (Whitmore [40])

In this paper, the author puts forward the inverse gaussian distribution as a description of length-of-hospital stay for certain types of illness and, particularly, for psychiatric illnesses. As an illustration, he compares the actual cumulative percentage of a cohort of 2,311 schizophrenic patients released from hospital at different times with the corresponding percentage based on a fitted inverse gaussian distribution. Although there is a significant departure of the observed percentage from that predicted by the fitted distribution, the fit is still quite good considering the very large sample size involved in the study and the difficulty which is experienced generally in finding acceptable theoretical models for psychiatric hospital-stay data.

Vol. 4
1976

His underlying conceptual model is again one of Brownian drift towards an absorbing barrier. In this case, at the moment of hospital admission, the patient's state of health is considered to be at level a. As time passes, the patient's health varies randomly but, nevertheless, it tends to drift towards health level b. At this level of health, the patient is considered to be sufficiently well for discharge. The first-passage time to this absorbing barrier is the patient's length-of-hospital stay.

Whitmore advances the inverse gaussian model not only as a *descriptive* model of the length-of-stay process but also as a *normative* model of the process. In other words, for certain illnesses the underlying natural process might reasonably be conjectured to be inverse gaussian. Therefore, observed systematic departures from this distribution might be viewed as artifacts of patient management techniques or limitations of the data. The detection of such departures would lead, generally, to the initiation of some remedial action such as an investigation of clinical practice, institution of better administrative controls, and so on.

In an earlier paper [39], Whitmore and Neufeldt also advance the inverse gaussian distribution as a potentially useful model for describing the length of

stay of a psychiatric patient in various patient states including those of inpatient, outpatient and nonpatient.

Realization time for fixed security price changes

A number of authors have advanced the idea that, in a reasonably efficient securities market, security prices will follow a random walk over time. (See, for example, the works reproduced in [3].) The idea has been refined considerably in recent years (cf. [13]) and much theoretical and empirical work has been done on it. The most popular version of the theory is the following. Let $P(0)$ denote the initial price of a security and let $P(t)$ denote its price at some subsequent time t; then $R(t) = \ln[P(t)/P(0)]$ for $t > 0$ is a Wiener process.

For certain financial securities such as stock options, warrants or convertible instruments, there is some special price P^* which is critical for the investor. For example, a bond which is convertible into common stock at a fixed ratio is liable to be called anytime after its conversion value exceeds the call price.[4] In these cases, interest often will lie in the length of time until the price first reaches P^*. In terms of $R(t)$, this is the first passage time for the process starting at 0 with an absorbing barrier at $R^* = \ln[P^*/P(0)]$. When $R(t)$ is a Wiener process with positive drift then this length of time has an inverse gaussian distribution. In the convertible bond case for instance, the time required for the conversion value to reach the call price for the first time is an inverse gaussian random variate. Thus, the inverse gaussian distribution can provide a theoretical model for such things as a valuation formula for convertible bonds, an empirical investigation of bond conversion times, and the like.

Queues and inventory systems

In [36], Wasan derives theoretical results for waiting-line systems with a single server; first, for the case of Poisson arrivals and inverse gaussian service times and, second, for the case of independent inverse gaussian interarrival times and exponential service times. He does not discuss specific case applications but it is not difficult to imagine a number of practical settings in which such waiting-line systems would be of interest. As one example, consider a mental health facility where patients arrive randomly (i.e. in a Poisson fashion) and have inverse gaussian lengths of stay as discussed earlier. If m is the patient capacity of the facility, then an m server system with Poisson arrivals and inverse gaussian service times is of interest in this case.

Closely related to waiting-line applications are applications of the inverse gaussian to inventory and storage problems. Wasan [36, pp. 156–163] describes an application due to Hasofer [6] and Gani and Prabhu [4] which involves water storage in a reservoir of infinite capacity. The application assumes that input to a water reservoir during a fixed time period is an inverse gaussian

[4] Actually, the conversion value usually will be some amount above the call price before an issue is called.

variate while water is released from the reservoir at a constant rate except when it is empty. A number of theoretical results are obtained such as the probability that the reservoir will be empty, transient behaviour of the reservoir contents, etc.

Whitt [42] and Loulou [12] have studied the asymptotic behaviour of waiting time and queue length in congested queuing systems and have found that these random variables are approximated well by normal distributions. Based on their results, it might be conjectured that the first passage time of the queue length or waiting time to some large capacity constraint or waiting threshold will be approximately inverse gaussian in a congested system. If this is the case, the inverse gaussian will be useful for establishing design parameters for systems which are subject to heavy traffic.

ADDITIONAL APPLICATIONS AND CONCLUDING REMARKS

Vol. 4
1976

The foregoing discussion of applications of the inverse gaussian distribution has been limited to those that already have received the attention of researchers or are on the threshold of realization. There are, however, numerous other areas where the distribution may find fruitful application. It is impossible to catalogue all of these but a few can be mentioned in closing. Business survival times or, more generally, the survival times of economic units in competitive or gaming situations, might be described by an inverse gaussian model. More precisely, if an entity with finite resources engages in a sequence of independent plays of an unfair game against an opponent with infinite resources, then the length of time until bankruptcy will have an approximate inverse gaussian distribution under fairly general conditions. Another area which offers attractive potential for inverse gaussian models is employment. Preliminary investigations suggest that periods of employment and unemployment may follow an inverse gaussian distribution. In fact, the duration of search activity, interpreted in a general way, might be explained by the model. In part, this is suggested by the occurrence of the inverse gaussian distribution in sequential sampling theory (see [33, p. 193]). The observed lengths of time required to complete various activities or projects might be explained by the inverse gaussian distribution under certain well defined conditions. If this is the case, then the distribution should prove valuable in project evaluation and review procedures. As a final application, it is worth mentioning the potential usefulness of the inverse gaussian model in describing the distribution of a parameter value in a population. Holla [7], for instance, uses the inverse gaussian distribution precisely in this manner to describe individual variation in accident proneness and to derive an explanatory model for accident data. The same model might be applied with

very little modification to susceptibility to disease and illness, and hence will be of interest to individuals concerned with the management of health systems.

The range of potential applications of the inverse gaussian distribution appears to be very wide. As mentioned earlier, the appropriateness of the model in any problem setting is determined largely by whether or not the diffusion process which characterizes the distribution is relevant to the problem at hand. Needless to say, the characterization also assists in reverse—activities which cannot be modelled by this diffusion process probably will not be explained very well by the inverse gaussian model.

REFERENCES

1. CHHIKARA RS and FOLKS JL (1974) Estimation of the inverse gaussian distribution function. *J. Am. Statist. Ass.* **69**(345), 250–254.
2. COX DR and MILLER HD (1965) *The Theory of Stochastic Processes.* John Wiley, New York.
3. COOTNER PH (Ed.) (1964) *The Random Character of Stock Market Prices.* M.I.T. Press, Cambridge, MA.
4. GANI J and PRABHU NV (1963) A storage model with continuous infinitely divisible inputs. *Proc. Camb. Phil. Soc.* **59**, 417–430.
5. HADWIGER H (1940) Eine analytische reproduktionsfunktion fur biologische gesamtheiten. *Skandinavisk Aktuarietidskrift* **23**, 101–113.
6. HASOFER AM (1964) A dam with inverse gaussian input. *Proc. Camb. Phil. Soc.* **60**, 931–933.
7. HOLLA MS (1966) On a Poisson-inverse gaussian distribution. *Metrika* **11**, 115–121.
8. JOHNSON NL and KOTZ S (1970) *Distributions in Statistics: Continuous Univariate Distributions*—2, Houghton Mifflin Co., Boston.
9. KHATRI CG (1962) A characterization of the inverse gaussian distribution. *Ann. Math. Statist.* **33**, 800–803.
10. KRAPIVIN VF (1965) *Tables of Wald's Distribution.* Nauka, Moscow. (In Russian).
11. LANCASTER T (1972) A stochastic model for the duration of a strike. *J. R. Statist. Soc.* (Ser. A), **135**, 257–271.
12. LOULOU R (1973) Multi-channel queues in heavy traffic. *Journal of Applied Probability* **10**, 769–777.
13. MERTON RC (1974) On the pricing of corporate debt: the risk structure of interest rates. *J. Finance* **24**(2), 449–470.
14. NADAS A (1973) Best tests for zero drift based on first passage times in Brownian motion. *Technometrics* **15**(1), 125–132.
15. PATEL RC (1965) Estimates of parameters of truncated inverse gaussian distribution. *Ann. Inst. Statist. Math.* (Tokyo) **17**, 29–33.
16. RAPOPORT A (1965) A note on equivalent mathematical models. *Bull. Math. Biophys.* **27**, 161–175.
17. ROY LK (1967) Some properties and applications of the inverse gaussian distribution. MSc Thesis submitted to Graduate School, Queen's University, Kingston, Ontario.
18. ROY LK and WASAN MT (1968) Properties of the time distribution of standard Brownian motion. *Revista Trabajos de Estadistica*, 67–82.
19. ROY LK and WASAN MT (1968) The first passage time distribution of Brownian motion with positive drift. *Mathematical Biosciences* **3**, 191–204.
20. ROY LK and WASAN MT (1969) A characterization of inverse gaussian distribution. *Sankhya*, March issue.
21. SAUNDERS SC (1974) A family of random variables closed under reciprocation. *J. Am. Statist. Ass.* **69**(346), 533–539.

22. SCHRODINGER E (1915) Zur theorie der fall-und steigversuchean teilchan mit Brownscher bewegung. *Phys. Z.* **16**, 289–295.
23. SESHADRI V and SHUSTER JJ (1974) Exact tests for zero drift based on first passage times in Brownian motion. *Techometrics* **16**(1), 133–134.
24. SHUSTER JJ (1968) A note on the inverse gaussian distribution function. *J. Am. Statist. Ass.* **63**, 1514–16.
25. SHUSTER JJ (1969) Properties of the inverse gaussian distribution. Unpublished doctoral thesis, McGill University, Montreal.
26. SHUSTER JJ and MIURA C (1972) Two-way analysis of reciprocals. *Biometrika* **59**, 478–481.
27. TWEEDIE MCK (1941) A mathematical investigation of some electrophoretic measurements on colloids. Unpublished thesis for MSc degree, University of Reading, England.
28. TWEEDIE MCK (1945) Inverse statistical variates. *Nature* **155**, 453.
29. TWEEDIE MCK (1947) Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc. Camb. Phil. Soc.* **43**, 41–49.
30. TWEEDIE MCK (1956) Some statistical properties of inverse gaussian distributions. *Va. J. Sci.* **7** (New Series), 160–165.
31. TWEEDIE MCK (1957) Statistical properties of inverse gaussian distributions. I. *Ann. Math. Statist.* **28**, 362–377.
32. TWEEDIE MCK (1957) Statistical properties of inverse gaussian distributions. II. *Ann. Math. Statist.* **28**, 696–705.
33. WALD A (1947) *Sequential Analysis*, John Wiley, New York.
34. WASAN MT and ROY LK (1967) Tables of inverse gaussian probabilities. (Abstract), *Ann. Math. Statist.* **38**, 299.
35. WASAN MT (1968) On the inverse gaussian process. *Skandinavisk Aktuarietidskrift*, Part 1–2, 69–96.
36. WASAN MT (1969) First passage time distribution of Brownian motion with positive drift (inverse gaussian distribution). *Queen's Papers in Pure and Applied Mathematics*, No. 19, Queen's University, Kingston, Ontario.
37. WASAN MT (1969) Sufficient conditions for a first passage time process to be that of Brownian motion. *Journal of Applied Probability* **6**(1), 218–223.
38. WASAN MT and BUCKHOLTZ PG (1973) Differential representation of a bivariate inverse gaussian process. *Journal of Multivariate Analysis* **3**, 243–247.
39. WHITMORE GA and NEUFELDT AH (1970) An application of statistical models in mental health research. *Bull. Math. Biophys.* **32**, 563–579.
40. WHITMORE GA (1974) The length-of-stay distributions of hospitalized psychiatric patients. Unpublished paper, Faculty of Management, McGill University, Montreal.
41. WHITMORE GA and YALOVSKY M (1975) Cumulative probability tables for the inverse gaussian distribution. Unpublished manuscript, Faculty of Management, McGill University, Montreal.
42. WHITT W (1970) Multiple channel queues and heavy traffic. *Advances in Applied Probability* **2**, 355–369.
43. ZIGANGIROV KS (1962) Expression for the Wald distribution in terms of normal distribution. *Radio Engng Electron. Phys.* **7**, 145–148.
44. ZIGANGIROV KS (1962) Representation of Wald's distribution by means of a normal distribution. *Radiotekh. Elektron.* **7**, 164–166. (In Russian).

ADDRESS FOR CORRESPONDENCE: *Professor GA Whitmore, Faculty of Management, McGill University, Montreal, Quebec H3A IG5, Canada.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Memoranda

An Heuristic for Sampling Rare Populations

THERE IS a vast literature of interest to management scientists and to managers on the problems of sampling inspection and quality control. However, one important problem which has received very little attention deals with the sampling inspection of lots having a small defective content where the manager would like either to estimate the true percentage defective or to 'catch' as many defectives as possible for a given sampling effort. This paper is a preliminary report on the performance of a novel heuristic for the problem of detecting as many defectives as possible.

Vol. 4
1976

THE PROBLEM

Consider the situation where a process yields items for which the percentage defective is low, say no greater than 2 per cent. Suppose also that management wishes to detect a significant number of the defective items for analysis of the nature and possible causes of the defects. Suppose further that the inspection of items is very costly (perhaps even destructive) so that the number of items inspected must be kept as low as possible.

Finally, suppose that there exists some basis for grouping the items into clusters such that the quality is likely to vary from cluster to cluster. This last assumption would be reasonable where a lot consists of items from different production lines or from different production periods, or where items used sub-assemblies from different manufacturers. This characteristic that the clusters have different levels of quality is important to the efficiency of the algorithm to be presented. It is not necessary that one be able to rate the clusters *a priori*, but simply that there be reason to believe they might differ in quality.

The problem posed is to detect as many defective items as possible using a fixed number of inspections. An inspection consists of the examination of an item which has been selected, by some rule, from the lot.

The proposed solution to the problem posed above had its genesis in a paper by Sudman [1] dealing with sampling from very rare human populations. Sudman briefly discussed the problem of selecting from large data files (e.g. Census tapes) the records corresponding to very rare individuals. Our heuristic is a further development of his concepts with the application in the quality control situation.

The steps in the algorithm are:

1. Group the items in the lot into a number of clusters, K.
2. For each cluster carry out a sequential probability ratio test of the hypothesis that the cluster is 'rich' in defectives relative to other clusters. Operationally, where the lot percentage defective is, say, 2 per cent, a cluster having a defective rate of, say, 3 per cent or higher will be called 'rich'.
3. For each 'rich' cluster, estimate the cluster percentage of defectives.
4. Group the 'rich' clusters into strata on the basis of their estimated defective rates so that those clusters estimated to have similar rates fall in the same stratum.
5. Sample disproportionately from the resulting strata (using, say, Neyman optional allocation) concentrating the effort on those strata most likely to contain defectives. The sequential probability ratio tests in (2) have provided

the necessary information on stratum variability to make this optimal allocation.

PRELIMINARY TESTS AND RESULTS

As a first step in evaluating the efficiency of the heuristic relative to other sampling schemes, a small simulation was conducted. The three methods compared were:

1. Simple random sampling of the entire lot.
2. Simple random sampling after 'poor' clusters have been eliminated by sequential testing.
3. The subject heuristic.

Simple random sampling was included largely as a validity check on the simulation, because the mean number of defectives which will be found using SRS is exactly the lot percentage defective. The second method gives some measure of the gains achieved due to the elimination of 'poor' clusters. Finally, the simulation of the heuristic gives a measure of the additional gains due to the stratification of the 'rich' clusters and the disproportionate sampling of these clusters.

Because the relative merits of the methods will depend on various parameters of the problem, the following parameters were considered:

(i) the number of clusters in the lot. (6, 10)

(ii) the lot percentage defective (1%, 2%)

(iii) the number of strata used in step 4. (2, 3)

(iv) the SPRT parameters (set *A*, set *B*).

The two sets of sequential probability ratio test parameters used were $A = (0{\cdot}05, 0{\cdot}1, 0{\cdot}0025, 0{\cdot}02)$, and $B = (0{\cdot}05, 0{\cdot}1, 0{\cdot}005, 0{\cdot}02)$ where the SPRT is defined by the 4-tuple $(\alpha, \beta, P_1, P_2)$. For each parameter two values were chosen and a 2^4 factorial experiment carried out in order to assess the effects of the various possible combinations of the parameter values. Of course the actual distribution of the defective items (the number in each cluster) plays a central role in determining the relative merits of the three methods. The approach was to distribute the defectives so that there were some 'very rich', some 'rich' and some 'very poor' clusters. A population was then constructed having the desired number of clusters and the defectives randomly distributed within clusters subject to the above guideline.

Table 1 presents the results of the simulation study, averaged over the $2^3 = 8$ values of the parameters other than the percentage defective and over 10 replications for each parameter set. The estimated standard error of each figure is $\pm$ 2 per cent. We note immediately that method I (simple random sampling) gives values very close to 1 per cent and 2 per cent, as expected, and that methods 2 and 3 represent improvements. As can be seen, the relative improvement from method I to method 2 is approximately 8·4 per cent and the increase from method 2 to method 3 is 6·8 per cent, with the overall improvement of the present heuristic over simple random sampling at 15·7 per cent.

TABLE 1. MEAN NUMBER OF DEFECTIVES FOUND PER INSPECTION

Sampling method	Lot % Defective 1%	2%
1	0·00993	0·0201
2	0·01080	0·0217
3	0·01152	0·0232

Further analysis of the results, especially with respect to the various parameters indicates the following:

The heuristic shows greatest promise for improvement in conditions of:

(i) A low percent defective;
(ii) Set *A* of the SPRT parameters;
(iii) A small number of strata, and
(iv) A large number of clusters.

We should reiterate that these results are preliminary and in need of further study. Whether the almost 16 per cent improvement in locating defectives is meaningful in a

given situation will depend on the economics of the situation, especially with respect to the cost function of the sampling scheme. The heuristic is certainly more complicated than SRS and we are presently studying the implications of reasonable cost functions.

ACKNOWLEDGEMENT

The authors acknowledge the support of the National Research Council of Canada, Grant Number 040226.

REFERENCES

1. SUDMAN S On sampling of very rare human populations. *J. Am. Stats. Ass.* **67** (338), 335–339.

MOHSEN ANVARI
ERIC N WEST
Concordia University
Sir George Williams Campus
Montreal, Quebec, Canada

(*November 1975*)

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Management Decision Rules in the Marketing of Electricity

A RECENT trend in marketing modeling is to replicate what managers do in practice. Parts of Kotler [5] and Montgomery and Urban [6] present examples of an approach which is based on determining the causal relationship of the factors used by executives in making marketing resource allocation decisions. Bailey and Malone [1] develop a production-oriented analysis which, with modifications, can be transformed into a model that is consistent with this framework. This note develops and analyzes this modified Bailey–Malone model, contrasts it with a traditional marketing resource allocation model, and briefly describes the implications for the electric utility firm and its customers. Data for the study comes from federal [2,3,4] and internal company sources [7,8].

BASIC MODELS

A variant of the Bailey–Malone [1, p. 135] regulated revenue maximizing model is consistent with management judgement, in which the electric utility firm (e.g. in Southern California), in the short run, seeks to:

maximize $R = P_1Q_1 + P_2Q_2$
subject to

$$[R - C - A - F - iK]/K = (r - i), \quad (1)$$

where

R = revenue;
P_1 = sector 1 (residential and commercial) market price;
P_2 = sector 2 (industrial market) price;
Q_1 = sector 1 demand $= q_{10}P_1^{q_{11}}A^{q_{12}}Y^{q_{13}}$

Q_2 = sector 2 demand $= q_{20}P_2^{q_{21}}V^{q_{22}}$,

where

q_{10}, q_{20} = constants of demand;
q_{11}, q_{12}, q_{13} = sector 1; constant demand response elasticities;
q_{21}, q_{22} = sector 2 constant demand; response elasticities:

K = capital investment (fixed in the short run);
F = fixed short run costs (exclusive of capital costs);
r = California Public Utility Commission determined 'fair' rate of return on investment;
i = firm's cost of capital;
A = discretionary promotion expenditures;
C = total variable costs of supplying electricity;
= various costs of generation, plus customer services (S), plus distribution (D), plus transmission (W) of electricity; the variable cost function C is
Y = household income, as a proxy for the exogeneous growth in residential and commercial sector demand for electricity,
V = value added, as a proxy for the exogeneous growth in industrial sector demand for electricity.

An alternative more traditional assumption is that, in the short run, the electric utility firm seeks to:

maximize $\pi = R - C - F - ik$
subject to $R = (1+r)(C + A + F + ik)$ (2)

where

π = profit;
$Q_1 = a_{10}P_1^{a_{11}}A^{a_{12}}S^{a_{13}}D^{a_{14}}Y^{a_{15}}$
$Q_2 = a_{20}P_2^{a_{21}}W^{a_{22}}V^{a_{23}}$.

MODEL DIFFERENCES

Forming La Grangean expressions for (1) and (2) and maximizing reveals that the marketing decision rules of the traditional

model are similar to those of the heuristic model, with the following key differences:

(1) in the traditional model, expenditures on channels of distribution and customer services are "demand stimulating activities," so that their elasticities determine the proportions used in setting the levels of the marketing policies, but, in the heuristic model, there are no such determinations in that the activities are merely components of the costs of supplying electricity;

(2) in the traditional model, market sector prices are proportions of the 'fair' markup on the average costs of generating electricity, but, in the heuristic model, prices are proportions of the average fixed plus regulated capital costs and of the average costs of supply (generation, distribution, and service);

(3) in the traditional model, transmission expenditures are determined by a proportion of the cost of generating electricity, while customers services and distribution expenditures are each different proportions of promotion, but, in the heuristic model, expenditures on channels of distribution and customer services are merely components of the costs of supplying electricity.

RESULTS

An econometric analysis of the statistical forms of the decision rules of both models indicates that the traditional model seems to be more in accordance with the available evidence than does the heuristic model. The main conclusions implied by the analysis are:

(1) the electric utility behaves as a profit maximizer in the short run in setting its marketing policies, subject to the constraint that revenues must be a 'fair' markup of its costs (to offset short run cost increases);

(2) there are inelastic market sector demand responses in the short run to industrial sector price and to the other individual marketing activities) (promotion, services, and transmission);

(3) the inelastic market sector demand responses to income and value added indicates that, in the short run, electricity is economically a relative 'necessity', in that an increase in available funds to consumers is expected to lead to a less than proportional increase in the demand for electricity;

(4) the short run cost function of the firms in this study increase at a decreasing rate, since the sum of the output elasticities is >1 for both models studied;

(5) the evidence suggests that the executives in the Southern California electric utility industry, in fact, do not completely implement their marketing decision rules.

REFERENCES

1. Bailey EE and Malone JC (1970) Resource allocation and the regulated firm. *Bell Jl Econ. Mgmt Sci.*, 129.
2. Department of Commerce (1952–71) *Survey of Current Business*. U.S. Government Printing Office, Washington.
3. Federal Power Commission (1952–71) *Statistics of Electric Utilities in the United States—Privately Owned*. U.S. Government Printing Office, Washington.
4. Federal Power Commission (1952–71) *Steam Electric Plants: Construction Costs and Annual Production Expenses*. U.S. Government Printing Office, Washington.
5. Kotler P (1971) *Marketing Decision Making: A Model Building Approach*. Holt, Rinehart and Winston, New York.
6. Montgomery DB and Urban GL (1969) *Management Science in Marketing*. Prentice-Hall, Englewood Cliffs.
7. San Diego Gas and Electric Company (1952–71) *Statistical Supplement to the Annual Report*.
8. Southern California Edison Company (1952–71) *Statistical Supplement to the Annual Report*. Southern California Edison Company, Los Angeles.
9. Zellner Arnold (ed.) (1968) *Readings in Economic Statistics and Econometrics*. Little, Brown, Boston.

GA Forgionne
(*December 1975*)

Business Management Department
California State Polytechnic University
Pomona, California 91768, USA

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Food Preparation by Critical Path Methods

THROUGHOUT the US military system, large scale central food preparation facilities (CFPF) are replacing traditional company size units. This has been mandated by the available cost savings, improved service and quality uniformity [1]. One problem faced in the CFPF is the scheduling of food to be prepared. There exists a repeated menu cycle in which the particular foods and quantity are pre-determined; equipment is fixed. The bulk of the food is prepared and then frozen until the scheduled day of serving, at which time the food is heated and served. The basic question addressed here is the scheduling of the food preparation so as to be feasible relative to the constraints inherent in the facility and in the relevant food technological considerations. Once feasibility is achieved, such considerations as minimum time schedules are of interest.

As input to this scheduling problem, it was required to perform a critical path analysis on each food at the appropriate batch size. These data were used to generate a schedule of equipment usage. An example is illustrated for preparation of a large batch of baked lasagna.

Table 1 below presents for each activity, the time required, equipment required, the 'early start time (*ES*)', 'late finish time (*LF*)', and 'slack time (*S*)', the latter three derived via standard critical path methods. Following Table 1 is the critical path diagram.

We consider the results in Table 1 and

TABLE 1. ACTIVITIES, TIMES, EQUIPMENT

Activities		t	ES	LF	S	Equipment
A	coat steam kettle	5	0	5	0	kettle 1 (K_1)
B	add beef and heat	10	5	15	0	K_1, agitator
C	drain fat	5	15	20	0	K_1
D	open and crush tomatoes	10	0	20	10	*VCM*
E	add tomatoes to meat and spice	25	20	45	0	K_1, agitator
F	reduce heat and simmer	30	45	75	0	K_1
G	adjust volume with water	5	75	80	0	K_1
H	cool sauce in kettle	60	80	140	0	K_1, agitator
I	transfer to panning area	5	140	145	0	—
J	add water, salt, oil to kettle and boil	25	0	105	80	kettle 2, (K_2)
K	add noodles and cook	20	25	125	80	K_2, agitator
L	drain and rinse	10	45	135	80	K_2
M	add oil and mix	5	55	140	80	K_2
N	transfer to panning area	5	60	145	80	—
O	combine eggs and cheese	5	0	140	135	mixer
P	transfer to panning area	5	5	145	135	—
Q	transfer cheese to panning area	5	0	145	140	—
R	pan lasagna	30	145	175	0	—
S	heat oven to 325°F	30	0	175	145	oven
T	place lasagna in oven	10	175	185	0	oven
U	bake for 30 minutes	30	185	215	0	oven
V	remove from oven	10	215	225	0	oven
W	transfer to freezer	0	225	225	0	—

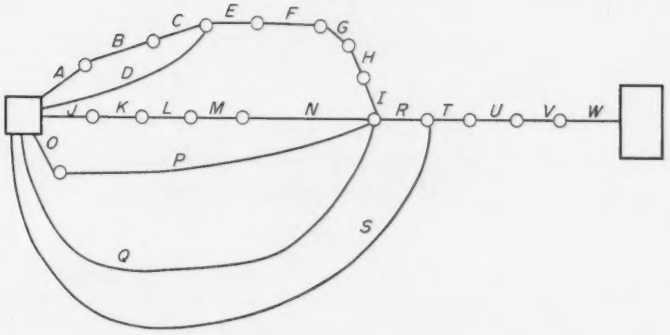


FIG. 1. *Critical path diagram.*

develop an 'Equipment usage summary.' consider the first kettle system (K_1). By noticing that all the activities requiring K_1 are on the critical path (no slack), the time requirement for K_1 is not flexible (assuming we wish to maintain the minimum time for total project completion), and is required during time interval 0–140. (We arbitrarily assume initially, for each food, a zero start time).

The second kettle system (K_2) is used during activities *J*, *K*, *L*, *M*, which comprise a total of 60 min. However, there is a lot of flexibility concerning which 60 min. By looking at the early start time of activity *J*, and the late finish time of activity *M*, we find that we need K_2 for any 60-min interval between 0–140.

The agitator summary is a little more complicated. When used with K_1, there is no choice and the required times are 5–15, 20–45, 80–140. However, when used with K_2 (activity *K*) it is required for 20 min, but it must be the 20 min selected for activity *K* once the 60 min for K_2 is selected.

The *VCM* is required for only 10 min, but it must be sometime in the 0–20 interval. The mixer is required for only 5 min and that 5 min can be anywhere in the 0–135 time interval. In both cases the time interval was determined by looking at the early start and late finish of the one activity using the particular piece of equipment.

Finally, we consider the oven. According to the analysis, it would be indicated that we require the oven from 175–225 with no choice (activities *T*, *U*, *V* on the critical path), and any 30-min period in the interval 0–175. However, this conclusion would be superficial. Clearly, if you heat the oven during any 30 min period before 145–175, it would not be available for (most) other products because the temperature is specific for the food under consideration. Thus we would conclude that the oven usage is essentially 145–225. (It is true that there may be some gain by producing at the same times, foods which require the same oven temperature. This fact was recognized but it was then decided not to consider it explicitly for the moment, since it was judged to be a factor to be built in later).

The purpose of generating these equipment usage summaries was to use them as input on the computer to generate via simulation a series of feasible production schedules. However, another purpose of the critical path analysis and equipment usage summary was to set up a complete program manual for the workers to follow during the preparation process. This is extremely useful, especially when considering that each day may find different people performing certain tasks in the CFPF.

REFERENCES

1. BUSTEAD R *et al.* (1972) A Proposed Modern Food Service System for Fort Lewis, *US Army Natick Laboratories Technical Report*, 73–10.

Dr PD BERGER
(*December 1975*)

School of Management
Boston University
212 Bay State Road
Boston
Massachusetts 00215
USA

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Some Numerical Equilibrium Results for the *GI/D/1*: (∞/*FIFO*) Queueing Model

THIS NOTE gives some numerical equilibrium results for the average queueing time of an arriving customer (W) and the probability of non zero queueing time (P) in models of the general type $GI/D/1:(\infty/FIFO)$ as defined in [3].

Tables 1 and 2 correspond to the case where inter-arrival times have the Erlang distribution given by:

$$dA(u) = \frac{(m\lambda)^m u^{m-1} \exp(-m\lambda u)}{(m-1)!}\, du$$

with mean $1/\lambda$, variance $1/(m\lambda^2)$ and coefficient of variation $v_a = \sqrt{1/m}$. For this case, the analytical method of solution discussed in [2] was adopted and applied for $m = 1, 2, 3, 4, 6$ and 10.

Table 3 is related to the three distributions of inter-arrival times: simple triangular (ST), rectangular (R) and double triangular (DT) with coefficients of variation equal to $\sqrt{1/2}$, $\sqrt{1/3}$ and $\sqrt{1/6}$ respectively. These distributions are the duals of the ones used in [1] in connection with service times.

For the three cases considered in Table 3 estimates of the average queueing time and of the probability of non zero queueing time were obtained by simulation using the SIMON language [3]. In each particular run 200,000 arrivals were simulated (the requirements in computer time being equivalent to approximately 8 min on a CDC 6400 computer) and variance reduction techniques were applied for the calculation of 95 per cent confidence limits on the simulated values.

A comparison between the numerical results included in this note and the ones presented in [1] in connection with the $D/G/1:(\infty/FIFO)$ model seems to suggest that, whereas the shape of the distribution of service times has little effect on the average queueing time in a $D/G/1:(\infty FIFO)$ system even for values of ρ well below unity, the average queueing time in a $GI/D/1:(\infty/FIFO)$ system depends not only on ρ and v_a but also on the shape of $A(u)$. From a practical viewpoint, therefore, unless ρ is close to unity the values in Table 1 may not be reliable estimates of the average queueing

TABLE 1. μW OBTAINED ANALYTICALLY FOR SOME $E_m/D/1:(\infty/FIFO)$ MODELS

ρ \ m	1	2	3	4	6	10
0·40	0·333333	0·103311	0·045011	0·022699	0·007168	0·001051
0·45	0·409091	0·136069	0·063416	0·034168	0·012310	0·002357
0·50	0·500000	0·176741	0·087202	0·049652	0·019938	0·004742
0·55	0·611111	0·227812	0·118032	0·070422	0·030956	0·008792
0·60	0·750000	0·293036	0·158409	0·098376	0·046674	0·015352
0·65	0·928571	0·378337	0·212283	0·136486	0·069113	0·025694
0·70	1·166667	0·493610	0·286247	0·189708	0·101603	0·041872
0·75	1·500000	0·656694	0·392188	0·266961	0·150115	0·067519
0·80	2·000000	0·903284	0·553900	0·386095	0·226565	0·109828
0·85	2·833333	1·316698	0·826924	0·588766	0·358736	0·185505
0·90	4·500000	2·146923	1·377912	0·999966	0·629966	0·344582
0·95	9·500000	4·643948	3·040186	2·244684	1·456922	0·837083

TABLE 2. P OBTAINED ANALYTICALLY FOR SOME $E_m/D/1:(\infty/FIFO)$ MODELS

ρ \ m	1	2	3	4	6	10
0·40	0·600000	0·220821	0·133799	0·085037	0·037146	0·008228
0·45	0·550000	0·270292	0·176573	0·120496	0·060363	0·017492
0·50	0·500000	0·323259	0·225273	0·163294	0·091732	0·033148
0·55	0·450000	0·379466	0·279748	0·213590	0·132233	0·057409
0·60	0·400000	0·438685	0·339830	0·271464	0·182692	0·092552
0·65	0·350000	0·500714	0·405337	0·336930	0·243788	0·140814
0·70	0·300000	0·565367	0·476085	0·409948	0·316062	0·204322
0·75	0·250000	0·632479	0·551884	0·490437	0·399928	0·285036
0·80	0·200000	0·701898	0·632542	0·578282	0·495683	0·384717
0·85	0·150000	0·773484	0·717871	0·673338	0·603517	0·504897
0·90	0·100000	0·847108	0·807681	0·775437	0·723525	0·646861
0·95	0·050000	0·922650	0·901785	0·884392	0·855711	0·811641

TABLE 3. μW AND P OBTAINED BY SIMULATION FOR SOME $GI/D/1:(\infty/FIFO)$ MODELS

Model description			Results of simulation runs					
			μW			P		
A(u)	v^2_a	ρ	lower 95% conf. limit	sim/ed	upper 95% conf. limit	lower 95% conf. limit	sim/ed	upper 95% conf. limit
ST	1/2	0·60	0·4160	0·4212	0·4264	0·4660	0·4674	0·4688
ST	1/2	0·70	0·6214	0·6330	0·6447	0·5710	0·5730	0·5750
ST	1/2	0·80	1·0325	1·0539	1·0754	0·6925	0·6939	0·6953
ST	1/2	0·85	1·4371	1·4688	1·5006	0·7596	0·7613	0·7630
ST	1/2	0·90	2·2370	2·2948	2·3527	0·8321	0·8341	0·8361
R	1/3	0·70	0·4225	0·4290	0·4356	0·4846	0·4867	0·4887
R	1/3	0·85	0·9617	0·9855	1·0093	0·6933	0·6947	0·6962
R	1/3	0·90	1·4965	1·5324	1·5684	0·7812	0·7830	0·7847
DT	1/6	0·85	0·4113	0·4213	0·4313	0·5791	0·5818	0·5844
DT	1/6	0·90	0·6796	0·6968	0·7141	0·6975	0·6995	0·7015
DT	1/6	0·95	1·4651	1·5149	1·5647	0·8361	0·8381	0·8401

time in a $GI/D/1:(\infty/FIFO)$ system for distributions of inter-arrival times other than the Erlang.

REFERENCES

1. COSMETATOS GP (1976) Some numerical equilibrium results for the $D/G/1:(\infty/FIFO)$ queueing model. *Omega* 4(1) 100–102.
2. COX DR and SMITH WL (1961) *Queues*. Methuen, London.
3. LEE AM (1966) *Applied Queueing Theory*. St. Martin's Press, Montreal, Canada.

GEORGE P COSMETATOS
(December 1975)

Department of Management Science
Imperial College of Science and Technology
Exhibition Road
London SW 7 2 BX, UK

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

Estimating Learning Time for Repetitive Tasks

THE TIME constant learning curve model is defined by,

$$Y_t = Y_c + Y_f\,[1 - \exp(^{-t}/\tau)] \quad (1)$$

where Y_t is the output rate at time t. Y_c, Y_f and τ are the model parameters describing the improvement in performance, Y_c being the initial output rate and Y_f being the ultimate improvement in output rate. τ is the learning time constant and has a significance in shaping the learning curve. Equation (1) has been found to apply to a wide variety of industrial tasks.

It is obviously advantageous if management can estimate model parameters from a few early results since meaningful long term forecasts are possible once the trend line is established. A number of computerized estimation techniques have been reported in the literature, e.g. [2], but the purpose of this memorandum is to offer an abbreviated description of a rapid hand method described in [1].

Using the notation $\dot{y}=dy/dt$, then from (1) we have $\dot{y}_t = (Y_c/\tau)\exp(-t/\tau)$, and

$$\dot{y}_{t+\Delta t} = \frac{Y_c}{\tau}\exp(-[t+\Delta t]/\tau)$$

hence

$$R = \frac{\dot{y}_t}{y_{t+\Delta t}} = \exp(^{\Delta t}/\tau). \quad (2)$$

It is clear from (2) that given two estimates of the tangent to the learning curve at points separated by Δt we have a simple method of estimating τ. In practice the data will be made available in discrete form as a time series defined as $\{y\} = \{y_0, y_1, y_2, \ldots y_n\}$ in which there are $(n+1)$ terms. For equispace data, a new series may be formed by difference to yield a new series which for convenience is defined by $\{x\} = \{(y_1-y_0), (y_2-y_1), \ldots (y_n-y_{n-1})\}$ so that $x_t = (y_{t+1}-y_t)$. At the mid-ordinate the slope is given approximately by the expression

$$\left(\frac{x_t}{\Delta t}\right)$$

and hence we have an estimate for R given by

$$\hat{R}_t = \left(\frac{x_{t-1}}{x_t}\right) \quad (3)$$

From (3) it is clear that an estimate of R may be made every 3 consecutive data points. Thus given a smoothed version of the industrial data the tangent to the curve may be calculated at a number of points, a number of estimates for R made, and an overall estimate made by averaging. Some smoothing must be used to damp out oscillations, otherwise wide swings in R will result. Once R has been estimated the future performance can be predicted in the following way.

Using (3), y_n may be re-written as

$$y_n = y_0 + x_0\left(\frac{1}{R^0} + \frac{1}{R^1} + \frac{1}{R^2} + \ldots \frac{1}{R^{n-1}}\right)$$

which on summation becomes,

$$y_n = y_0 + x_0\left[\frac{R^n-1}{R^{n-1}(R-1)}\right] \quad (4)$$

which can be used for forward projection when estimates $\hat{R}$ and $\hat{x}_0$ are available. Given $\hat{R}$, $\hat{x}_0$ is estimated by,

$$\hat{x}_0 = (y_n - y_0)\left[\frac{R^{n-1}(R-1)}{R^n-1}\right] \quad (5)$$

As an example, consider the industrial data of Table 1, when forecasts of performance on day 50 ahead are required using only data available up to day 25. Using the smoothed values, there are six points ($n=5$) in the data sequence, $\{y\} = \{19, 35, 47, 55, 61, 67\}$ giving $\{x\} = \{16, 12, 8, 6, 6\}$ and $\{\hat{R}\} = \{1{\cdot}33, 1{\cdot}50, 1{\cdot}33, 1{\cdot}00\}$. $\hat{R}_{AV}$ is thus 1·283 and using (5), $X_0 = 14{\cdot}96$. The prediction for day 50 (the 11th data point hence $n = 10$) is then, from (4)

$$y_{11} = 19 + 14{\cdot}96\left\{\frac{1{\cdot}283^{10}-1}{1{\cdot}283^{9}(1{\cdot}283-1)}\right\}$$

$$= 81{\cdot}08$$

Vol. 197

TABLE 1. INDUSTRIAL DATA AND FORECASTS

Day No.	D	5	10	15	20	25	40	50	60	65
Y_t Actual	19	32	46	54	50	70	76	76	71	88
Y_t Smoothed	19	35	47	55	61	67				
Y_t Forecast from Day 25							85	85+	85+	86

τ for this curve is given by (2). Using $\hat{R} = 1{\cdot}283$,

$$\tau = \frac{\Delta t}{\ln R} = \frac{5}{\ln\ 1{\cdot}283} = 20 \text{ days.}$$

$\hat{Y}_f$ may now be calculated from (1), since when $t = 5$ days, the time varying part is $\hat{x}_0$, hence

$$\hat{Y}_f = \frac{14{\cdot}96}{1 - \exp(-5/20)} = 67{\cdot}6$$

REFERENCES

1. HINDMARSH GW and TOWILL DR (1974) Theory and application of the time constant learning curve model. Paper WA16.14. ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico.
2. HITCHINGS B and TOWILL DR (1975) An error analysis of the time constant learning curve model. *Int. Jnl. Prod. Res.* **13** (2), 105–135.

GW HINDMARSH
DR TOWILL
(*November 1975*)

Industrial Engineering Unit
Department of Mechanical Engineering and Engineering Production
University of Wales Institute of Science and Technology
Cardiff
UK

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 2, 1976. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

The following are available at the Yvette Jacobson Library, Management Centre, University of Bradford, Emm Lane, Bradford, Yorkshire BD9 4JL, U.K.:

BARKER CR (1970) Factors in the Development of Industrial and Commercial Training with Particular Reference to the Behavioural Sciences and Management Training. MSc Thesis, University of Bradford.

BROWN PA (1973) The Use and Selection of Service Levels in the Control of Inventory for Marketing. MSc Thesis, University of Bradford.

BRUMMER A (1972) A Social Systems Approach to Marketing. MSc Thesis, University of Bradford.

BURKE MCP (1973) An Assembly Line Balance for the Telecommunication Industry—The 2000 Type Selector Mechanism. MSc Thesis, University of Bradford.

BURN RE (1972) Aspects of a Passenger Transport System. MSc Thesis, University of Bradford.

BUXTON G (1971) A New Approach to Depot Delivery Areas in Logistics Planning. MSc Thesis, University of Bradford.

COLLINS JK (1971) The Organisational and Educational Status of the Procurement Function: A Study in Selected Areas of British Industry. MSc Thesis, University of Bradford.

COOKE PNC (1970) A Market Taxonomy Utilising Cultural Anthropological Indicators. MSc Thesis, University of Bradford.

COOKE SF (1970) Recruitment and Selection of Salesmen within a Telecommunications Company. MSc Thesis, University of Bradford.

COWELL DW (1973) The Effectiveness of Marketing Courses. PhD Thesis, University of Bradford.

CREW MA (1971) The Theory of Peak Load Pricing and its Application. PhD Thesis, University of Bradford.

DOBBINS R (1971) Institutional Shareholders in the Equity Market. MSc Thesis, University of Bradford.

GUHA PS (1973) The Significance of Cultural Anthropology in Marketing Strategies. MSc Thesis, University of Bradford.

HALL DS (1972) The Determination of Optimum Shop Size: A Study of Electrical Retailing. MSc Thesis, University of Bradford.

HANRAHAN JM (1970) A Study of Attitudes Towards and Effectiveness of Direct Mail Advertising to Medical Practitioners in the United Kingdom. MSc Thesis, University of Bradford.

HARDY AG (1972) An Investigation into Some Areas of Industrial Relations Existing in Engineering. MSc Thesis, University of Bradford.

HINDLE MW (1970) An Investigation of Public Relations in the Pharmaceutical Industry. MSc Thesis, University of Bradford.

Vol. 4
1976

HOWDLE JM (1973) Acquisition Planning in the U.K. Carpet Industry. MSc Thesis, University of Bradford.

HOWLEY JC (1971) The Firm and Its Environment; An Empirical Systems Analysis of the Impact of Its Environment Upon a Business Organisation. MSc Thesis, University of Bradford.

IRVING P (1970) Corporate Planning in Practice: A Study of the Development of Organised Planning in Major U.K. Companies, MSc Thesis, University of Bradford.

JACKSON KC (1970) The Lancashire Textile Industry Since 1950, A Case Study of Government Intervention in Private Industry. MSc Thesis, University of Bradford.

LIVERSIDGE FCA (1971) A Critical Analysis of the Investment Decision. MSc Thesis, University of Bradford

MCARTHUR AJ (1972) To Create and Examine New Ideas Concerning Product Markets for Middle Distillate Fuel Oils. MSc Thesis, University of Bradford.

MARTIN JC (1973) The Computer in Production and Operations Management. MSc, Thesis University of Bradford.

MORGAN JC (1972) Inter-organisational Distribution Channels: An Empirical Study. MSc Thesis, University of Bradford.

MORRIS S (1972) An Investigation into the Workload of a Salesman. MSc Thesis, University of Bradford.

NICHOLSON DJP (1972) An Initial Comparison of Two-Vehicle Scheduling Programs. MSc Thesis, University of Bradford.

OLDHAM KM (1973) Accounting Systems in Europe. MSc Thesis, University of Bradford.

OPENSHAW DK (1972) The Feasibility of Developing a Linear Programming Model of Wool Blending. MSc Thesis, University of Bradford.

QUAYLES NJT (1971) Controlling a Company's Transport Operations. MSc Thesis, University of Bradford.

ROTHWELL AE (1970) Operational Research in Public Cleansing Systems: Some Applications of a Basic Model. MSc Thesis, University of Bradford.

The following are obtainable at the University of California Library, Berkeley, California 94720, *U.S.A.:*

CARLSON D (1972) Financial Analysis of Institutions of Higher Education for Policy Decisions. PhD Dissertation, University of California, Berkeley.

CURRY DJ (1973) An Empirical Test of Relations Between An Individual's Preference and Similarity Space of Brands. PhD Dissertation, University of California, Berkeley.

DAY DN JR (1972) A Multi-Stage Learning Model. PhD Dissertation, University of California, Berkeley.

FALTAS NC (1972) Methodological and Implementational Issues in the Design of International Economic Relations. PhD Dissertation, University of California, Berkeley.

FEUILLE P (1973) Police Union Power. PhD Dissertation, University of California, Berkeley.

FIORELLO MR (1972) Management and Design Tools for Document Retrieval Systems: A Method for Predicting Quantity Output. PhD Dissertation University of California, Berkeley.

GAYTON RJ (1973) A Critical Examination of Presently Accepted International Accounting Translation Techniques. PhD Dissertation University of California, Berkeley.

GOTCHER JW (1972) A Systems Approach to Policies Governing the Design of an Information Dissemination System for the Earth Resources Technology Satellites. PhD Dissertation, University of California, Berkeley.

JOHANSSON JK (1972) Returns to Scale in Advertising Media. PhD Dissertation, University of California, Berkeley.

KLEIN D (1973) Dynamic Behaviour of Adjustment Processes. PhD Dissertation, University of California, Berkeley.

KUHN AJ (1972) An Application of a System Control Model to Business History: The General Motors Corporation under Alfred Sloan, 1920 to 1935. PhD Dissertations, University of California, Berkeley.

LAMM R (1973) Black Union Leaders at the Local Level. PhD Dissertation, University of California, Berkeley.

MAC LACHLAN DL (1971) Market Responsive Demand Models: A Comparison of Variable Markov and Distributed-Lag Approaches. PhD Dissertations, University of California, Berkeley.

MEDOFF MH (1973) A Theoretical and Positive Analysis of Discrimination in the Market Place. PhD Dissertation, University of California, Berkeley.

MISTER W (1973) Temporal Additivity: A Problem of Accounting Measurement. PhD Dissertation, University of California, Berkeley.

MOLANDER EA (1972) The Idealogy of Aerospace Program Managers. PhD Dissertation, University of California, Berkeley.

MOORE JH (1973) Peak-Load Allocation of a Time-Shared Computer in a Quadratic Team. PhD Dissertation, University of California, Berkeley.

MORRIS JR (1972) The Decision to Purchase Information for Portfolio Choice. PhD Dissertation, University of California, Berkeley.

MONAKA I (1972) Organization and Market: Exploratory Study of Centralization and Decentralization. PhD Dissertation, University of California, Berkeley.

OHLSON JA (1972) Analysis of the Usefulness of Accounting Data for the Portfolio Decision: A Decision-Theoretic Approach. PhD Dissertation, University of California, Berkeley.

PORTWOOD CS (1972) Statistics and Their Application to Social Sciences Research. PhD Dissertation, University of California, Berkeley.

PURVES RA (1972) Interstate Migration, Demographic Trends and Regional Employment Growth: An Analysis of Recent Migration to California. PhD Dissertation, University of California, Berkeley.

SADAN S (1972) The Firm-Investor Interface: An Inquiry into the Role of Accounting Information. PhD Dissertation, University of California, Berkeley.

SIMKIN MG (1972) Mathematical Programming for the State Earmarking Process: Tax Policy in Theory and An Application for Highway Finance. PhD Dissertation, University of California, Berkeley.

SNOW CC (1972) Environmental Uncertainty and Organization Behaviour. PhD Dissertation, University of California, Berkeley.

SOLOMON S (1972) Determination of Satisfactory Scale for Data Processing facilities. PhD Dissertation, University of California, Berkeley.

STEBBINS MW (1972) The Social Science PhD in Corporate Industry. PhD Dissertation, University of California, Berkeley.

WALTERS KD (1972) Freedom of Speech in the Modern Corporation. PhD Dissertation, University of California, Berkeley.

The following are obtainable at the Joseph Regenstein Library, University of Chicago, 1100 *East 57th Street, Chicago, IL* 60637, *U.S.A.:*

BLOCK MW JR (1974) Member Commitment to Group Decisions: A Study of Individual and Group Determinants. PhD Thesis, University of Chicago.

GEISEL MS (1970) Comparing and Choosing among Parametric Models: A Bayesian Analysis with Macroeconomic Applications. PhD Thesis, University of Chicago.

Vol. 4
1976

GRAYSTON RW (1971) Deterrents in Auto Liability Insurance. PhD Thesis, University of Chicago.

HALPERN PJ (1971) Empirical Estimates of the Expected Economic Gains from Mergers and an Analysis of Their Division between Companies in the Merger. PhD Thesis, University of Chicago.

HOPWOOD AG (1971) An Accounting System and Managerial Behavior. PhD Thesis, University of Chicago.

JAFFE JF (1973) Security Regulation, Special Information, and Insider Trading. PhD Thesis, University of Chicago.

KNOWLES TW (1971) An Artificial Variable Elimination Method for Solving Block Diagonal Programming Problems. PhD Thesis, University of Chicago.

LAUB PM (1971) The Divided-Earnings Relationship: A Study of Quarterly Corporate Panel Data. PhD Thesis, University of Chicago.

VON BRIESEN NEUHAUSER D (1971) Specification of Procedures, Visibility of Consequences, and Organizational Efficiency: An Empirical Test of the Entrepreneurial Theory of Formal Organizations Using Data from Short-Term General Hospitals. PhD Thesis, University of Chicago.

NOURIE FJ (1970) A Theoretical Study of Cutting-Plane Methods for Integer Programming. PhD Thesis, University of Chicago.

OFFICER RR (1971) An Examination of the Time Series Behavior of the Market Factor of the New York Stock Exchange. PhD Thesis, University of Chicago.

PEKELMAN D (1971) Optimal Production Path with Adjustment Cost and Time-Dependent Price. PhD Thesis, University of Chicago.

POLLAY RW (1970) A Theoretical Investigation into Factors Affecting Decision Completion Time. PhD Thesis, University of Chicago.

RICHARDSON WC (1971) Ambulatory Use of Physicians' Services in Response to Illness: Episodes in a Low-Income Neighborhood. PhD Thesis, University of Chicago.

RUBIN DS (1970) The Neighboring Vertex Cut and Other Cuts Derived with Gomory's Asymptotic Algorithm. PhD Thesis, University of Chicago.

SCHOLES MS (1970) A Test of the Competitive Market Hypothesis: The Market for New Issues and Secondary Offerings. PhD Thesis, University of Chicago.

SCOTT WR (1973) A Bayesian Approach to the Accounting Problems of Asset Valuation and Audit Size. PhD Thesis, University of Chicago.

SEXTON DE JR (1970) Do Blacks Pay More? PhD Thesis, University of Chicago.

SHORTELL SM (1972) Physician Referral Behavior: A Test and Further Specification of Exchange Theory. PhD Thesis, University of Chicago.

TAYLOR DG (1973) The Effects of Monetary Growth on Interest Rates. PhD Thesis, University of Chicago.

THOWSEN GT (1973) A Dynamic Nonstationary Inventory Problem for a Price/Quantity Setting Firm. PhD Thesis, University of Chicago.

TRAVERS GF (1973) The Structure of Minority Occupational Position. PhD Thesis, University of Chicago.

WILLIAMSON RW JR (1971) The Use of the Accounting Rate of Return as an Estimate of the Internal Rate in Financial and Economic Analyses. PhD Thesis, University of Chicago.

The following are obtainable at the Library, Cranfield Institute of Technology, Cranfield, Bedford MK43 OAL, U.K.:

CLARKE CD (1970) Project Management. MBA Thesis, Cranfield Institute of Technology.

COLEMAN CD (1970) An Investigation into the Determinants of Executive Success. MBA Thesis, Cranfield Institute of Technology.

Vol. 4
1976

DAVIS JB (1970) Corporate Strategy, Archibald Kenrick & Sons Ltd. MBA Thesis, Cranfield Institute of Technology.

DOUGLASS TJ (1971) The Executive Aircraft as a Business Tool. MBA Thesis, Cranfield Institute of Technology.

Fairweather R (1971) Group Delivery Performance Study. MBA Thesis, Cranfield Institute of Technology.

FEELY SP (1971) A Survey of the United Kindom Charity Market. MBA Thesis, Cranfield Institute of Technology.

GADGIL SG (1971) A Marketing Information System for a Squares Department. MBA Thesis, Cranfield Institute of Technology.

GOMM J (1971) The Assessment of Management Potential. MBA Thesis, Cranfield Institute of Technology.

GRAY JA (1972) Commercialisation of New Materials with Particular Reference to Carbon Fibres. MBA Thesis, Cranfield Institute of Technology.

HAMILTON RI (1972) A Study of the Operations of a Garage Equipment Supplier in the U.K. Vehicle Servicing Market. MBA Thesis, Cranfield Institute of Technology.

HAY WG (1972) A Feasibility Study on the Establishment of Corporate Standards Covering the Selection Procedures for First Line Production Supervisors with the British Leyland Motor Corporation, Using the Assessment Centre Technique as a Possible Model. MBA Thesis, Cranfield Institute of Technology.

MASON TG (1972) Development of a Market and Corporate Strategy for the Leicester Shop Equipments Co. Ltd. MBA Thesis, Cranfield Institute of Technology.

NEATE PW (1972) A Survey of the Electronic Passive Sub-Component Market with a View to Acquisition. MBA Thesis, Cranfield Institute of Technology.

PAYTON TH (1972) A Model for Estimating Market Research Service Fees. MBA Thesis Cranfield Institute of Technology.

PRIOR GA (1970) The City Code on Takeovers and Mergers. MBA Thesis, Cranfield Institute of Technology.

REYNOLDS MC (1970) Agricultural Land Values Since 1945. MBA Thesis, Cranfield Institute of Technology.

ROBERTS TA (1970) Recommendations to a Sub-Contract Machining Company Resulting from a Survey of the Sub-Contract Machining Market. MBA Thesis, Cranfield Institute of Technology.

SKINNER RPB (1970) An Examination into the Role of the Area Manager of a Multiple Supermarket Company. MBA Thesis, Cranfield Institute of Technology.

SMITH PML (1971) The Problems Facing Very Small Companies with Special Reference to the Welsh Woolen Industry. MBA Thesis, Cranfield Institute of Technology.

TANNER BS (1971) The Development of Air Conditioning in the U.K. MBA Thesis, Cranfield Institute of Technology.

TERRY PM (1971) A Study of Fire Incidence in Towns. MBA Thesis, Cranfield Institute of Technology.

THOMPSON NP (1971) A Comparison of the Economics of Beef Production in the U.K. and the E.E.C. MBA Thesis, Cranfield Institute of Technology.

TRIPP RF (1971) An Evaluation of the Market for a Mechanically Propelled Vehicle. MBA Thesis, Cranfield Institute of Technology.

WHITE RC (1971) Intercompany Trading in an Engineering Group. MBA Thesis, Cranfield Institute of Technology.

WHITEHEAD TP (1971) An Investigation of the Tyre-Mounted Combined Excavator/Loader Market in the U.K. MBA Thesis, Cranfield Institute of Technology.

WRIGHT GR (1971) A Department Analysis, with Special Reference to Profitability, in a Department Store. MBA Thesis, Cranfield Institute of Technology.

WRONSKI M (1973) The Impact of International Container Shipments on Corporate Physical Distribution Management. PhD Thesis, Cranfield Institute of Technology.

YEALLAND JS (1971) A Study of BOC's Food Import–Export Business in Western Europe with Particular Regard to the Effect of the Entry of Great Britain and Eire into Membership of the European Economic Community. MBA Thesis, Cranfield Institute of Technology.

The following are available at the Lyon Playfair Library, Imperial College of Science and Technology, Exhibition Road, London SW7 2BX, U.K.:

AHMAD TU (1973) An Examination of Production Scheduling—Theory and Practice. MSc Thesis, Imperial College, London.

ALEEM I (1973) Building a Corporate Financial Model. MSc Thesis, Imperial College, London.

BRAILSFORD A (1971) Resource Allocation in Network Analysis. MSc Thesis, Imperial College, London.

BRICE JD (1971) The Market's Judgement of Firms Raising New Money. MSc Thesis, Imperial College, London.

CHAN KS (1972) A Stock Control Study at a Distributor. MSc Thesis, Imperial College, London.

CHAPMAN IC (1972) Capital Investment Appraisal by Computer Simulation Model. MSc Thesis, Imperial College. London.

CHATTERJI SN (1972) The Process of Innovation. MSc Thesis, Imperial College, London.

CHEE PCF (1971) The Design of Fixed Delivery Areas: A Case Study. MSc Thesis, Imperial College, London.

CLINCH A (1972) The Early Redemption of Commercial Loans. MSc Thesis, Imperial College, London.

COHEN M (1971) Two Algorithms which Find the Maximum Capacity–Reliability Route Through A Network. MSc Thesis, Imperial College, London.

DICKEY RW (1971) The Evaluation of Potential Bank Branches. MSc Thesis, Imperial College, London.

EADY GT (1972) An Investigation into the Present and Future of Some Aspects of Automation in the Hotel Industry. MSc Thesis, Imperial College.

FRAGOULIS NS (1973) A Demand Model for an n Commodity Case. MSc Thesis, Imperial College, London.

FRANK MT (1973) The Economics of Scrap Metal Recovery. MSc Thesis, Imperial College, London.

HARIHARAN A (1973) A Decision-Process Model of the Investment Function in the Manufacturing Sector of the Macroeconomy. MSc Thesis, Imperial College, London.

JAMES PA (1971) A Comprehensive Bibliography and Abstracts in Industrial Scheduling. MSc Thesis, Imperial College, London.

JEYNES JP (1973) The Use of Linear Programming in Stock Evaluation. MSc Thesis, Imperial College, London.

Kaur S (1973) Developing a Forecasting Model for Clothing Demand. MSc Thesis, Imperial College, London.

KENDALL JW (1973) Enhancement and Use of Creativity in Business. MSc Thesis, Imperial College, London.

KOLB KH (1972) Automation in Share Registration. MSc Thesis, Imperial College, London.

LAWRENCE PA (1971) Accounts Payable for Financing Small Companies During Periods of Credit Restriction. MSc Thesis, Imperial College, London.

LEPAGE DE (1971) A Separable Programming Formulation for Scheduling Continuous Production. MSc Thesis, Imperial College, London.

MANN WR (1971) Operational Research and Urban Planning. MSc Thesis, Imperial College, London.

POON LC (1972) An Investigation into the Effect of a Productivity Agreement on Labour Productivity within a Company. MSc Thesis, Imperial College, London.

READ S (1972) Analysis of Retail Credit. MSc Thesis, Imperial College, London.

SAHGAL B (1972) Measures of Productivity and Efficiency—An Empirical Study. MSc Thesis, Imperial College, London.

SERGHIOU SS (1973) Robustness of Results in Job Shop Simulation. MSc Thesis, Imperial College, London.

YIP JK (1971) Inventory Levels in Multi-Echelon System. MSc Thesis, Imperial College, London.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

EDITORIAL

Equal Pay

SOMETIMES I cannot help having a sneaking suspicion that there is little new under the sun, that major issues with which we grapple today have been with us (in one form or another) before, that many so-called new ideas and theories about life and society are mere transmutations of old arguments. I certainly had this feeling the other day when I came across some essays, one written almost 50 years ago by Aldous Huxley [1], the others over a hundred years ago by John Ruskin [2]. The two essayists, renowned for their remarkable insights and persuasive styles, lived in very different eras, their writings naturally reflect the problems of their environment and the examples they cite sound dated, but there is almost a sense of timelessness about some of their discussions; one gets an uncanny feeling that, whatever we think of their conclusions, the arguments have remained almost unscathed and their perspective undimmed by the passage of decades. On the theme of equal pay they held very different views, and in this respect perhaps the only thing they had in common was their strong criticism of contemporary thinking.

Vol. 4
1976

Huxley's essay is largely concerned with the prospect of a communist revolution, starting with Marx's "celebrated and almost universally unread work", which postulated the continued existence of a proletariat, exploited and victimized at the hands of the capitalists, until a violent revolution overturns the system to found a communist state. But Huxley believes that Marx was wrong in assuming that the proletariat would otherwise continue to be exploited and therefore continue to exist. Changing tenets of capitalism involve teaching the proletariat how to be wasteful, how to consume, even how to mortgage their future for the sake of material well-being, and hence it is in the interest of capitalists "to pay well, because those who are paid well buy well, particularly when hypnotised by the incessant suggestions of modern advertising", and all this in order to stimulate production of all goods and services, even when this involves the use and possible extinction of scarce raw material resources. A prosperous proletariat, that is what Huxley believes modern capitalism wants, and such a proletariat is no longer oppressed and deprived (in the material sense), but becomes part of the bourgeoisie, a "transmogrified proletariat", no

more interested in revolution to achieve what they already have than other sections of the bourgeoisie. (It is interesting, however, to note that Huxley did not consider the possible effects of major economic disasters, such as the depression of the early thirties.)

The major reason for this change Huxley attributes to "the equalisation of income—that happy consummation from which Mr. Bernard Shaw expects all blessings automatically to flow", a process which has already eroded differentials and which in time ("by a gradual and more or less painless process of propaganda, pressure of public opinion, and finally legislation") will lead to essentially equal incomes for all, which, he argues, may in some ways be obviously good and in others bad ("it would mean, for example, the complete practical realisation of the democratic idea, and this in its turn would mean, almost inevitably, the apotheosis of the lowest human values and the rule, spiritual and material, of the worst men). But good or bad, the equalisation of income can no more touch the real sources of present discontent than could any other large-scale book-keeping operation". Huxley sees equal pay as a development that eliminates the danger of a communist revolution, because it becomes irrelevant, but the consequence is that the social system "makes life fundamentally unlivable for all" and he fears that the equalization of incomes will lead to a "great orgy of universal nihilism".

In retrospect we know that, whatever effect the prosperity of the proletariat has had in averting revolutions in many industrialized countries, it has not prevented the emergence of new communist regimes, nor the intensified arguments in some countries as to who should be ultimately in charge of the means of production and the government of industry. In this respect Huxley fails to acknowledge that conflicts in society are primarily about the distribution of power, and it is Ruskin who reminds us of the purpose of acquiring wealth: "What is really desired, under the name of riches, is, essentially, power over men; in its simplest sense the power of obtaining for our advantage the labour of servant, tradesman, and artist; in wider sense, authority of directing large masses of the nation to various ends . . . And this power of wealth of course is greater or less in direct proportion to the poverty of the men over whom it is exercised, and in inverse proportion to the number of persons who are as rich as ourselves, and who are ready to give the same price for an article of which the supply is limited". He adds, "the art of making yourself rich, in the ordinary mercantile economist's sense, is therefore equally and necessarily the art of keeping your neighbour poor".

Seen in this light, the quest for increased prosperity of the working class, the demand for higher wages, the arguments about the division of the national cake (as well as the profits of individual enterprises) are really manifestations of struggles for power. At first, when labour was plentiful and unorganized, the purpose of these struggles was to ensure that wages were sufficient to maintain minimum subsistence levels, to achieve material benefits for the exploited and

the victimized, to fight starvation and misery. These aims were real and immediate, there was no need to mask them, there was no insidious power struggle. But all the material advantages extracted from the employers have not quenched the appetite for more, and—as Huxley points out—this appetite is constantly nurtured by aggressive advertising to stimulate ever increasing demand. The arguments (in the developed countries) are not centred nowadays on getting enough to subsist. Having enough is no longer enough. The arguments are more blatantly related to power, to the desire not only to become part of a system but to control it, to the exercise of power in order to attain more power. If that is true, it follows that more prosperity for the proletariat need not necessarily mitigate violent conflict in industry. The corollary of this conclusion is that in a democratic society, where people are allowed to form alliances and pressure groups, conflict is unending, the resolution of one inevitably sows the seeds of another; conflict only changes in form and intensity, its manifestations change, so do the rules of the game and the various protagonists.

How, then, is one to judge the justification for, and the merits or otherwise, of equal pay? The arguments may be marshalled along two fronts: the first is concerned with the moral issue as to whether it is right for one human being or a group of individuals to consume more of the produce of this Earth than another, or for one nation to be more prosperous than another, and whether people should be rewarded according to their needs or according to their worth to the community (and if so, which community?). The second is concerned with outcomes, the possible effect on the system and its performance, the resultant sense of acquiescence or alienation of its members, the resilience that the system can build to protect itself from outside threats, and, superimposed on all these considerations, the projected time scale (in that it is possible all too often to attain certain objectives in the short term at the expense of the long term).

Although proponents of equal pay usually admit to a mixture of motives, it is clear that the two sets of moral and pragmatic arguments are quite separate and that any suggestion that they are or even must be compatible is blatantly erroneous. Thus, the proposition that equal pay is bound to be so beneficial not only to the individual but also to the system to which he belongs, that it will abate conflict and will therefore be embraced by all, is patently dubious; equal pay does not imply prosperity for all, prosperity (as we have seen) does not necessarily result in the abolition of conflict, industrial conflict arises not only from motives of material gain but (in industrialized societies) primarily for reasons of power struggle, and what benefits all is not necessarily accepted by all.

Ruskin was not so much concerned with the pragmatism of universal acceptability (at one point he states that "all endeavour to deduce rules of action from balance of expediency is in vain"). Although there were violent conflicts on the shop floor and many strikes in his day, the development of organised labour was still in its infancy and it is too much to expect of him

Vol. 4
1976

to have visualized the sophisticated twists and turns that future industrial turmoils came to develop. His castigation of political economists, whom he regarded as totally incapable of providing any solution to reconcile the opposing parties in industry, indicates that he was interested in measures that would avert "an explosive action of disaffection". In his essays he therefore considers two points: "The first—How far the rate of wages may be so regulated as not to vary with the demand for labour. The second—How far it is possible that bodies of workmen be engaged and maintained at such fixed rate of wages (whatever the state of the trade may be), without enlarging or diminishing their number, so as to give them permanent interest in the establishment with which they are connected". He is very critical of Mill's attempt to distinguish between direct and indirect work and his assertion that demand for goods is not demand for labour. He takes a dim view of Ricardo's discussion of wages: "Ricardo, with his usual inaccuracy, defines what he calls the 'natural rate of wages' as 'that which will maintain the labourer'. Maintain him! yes; but how?" (i.e. what quality of life?).

Ruskin totally rejects the prevailing economic proposition that manpower should be treated like a commodity and that its price should be determined by supply and demand, i.e. that it should equal the marginal productivity of labour (or the marginal revenue product) at full employment. He takes political economy to task for its "idea that an advantageous code of social action may be determined irrespectively of the influence of social affection", and goes on: " 'The social affections', says the economist, 'are accidental and disturbing elements in human nature; but avarice and the desire of progress are constant elements. Let us eliminate the inconstants, and, considering the human being as a covetous machine, examine by what laws of labour, purchase, and sale, the greatest accumulative result in wealth is obtainable. Those laws once determined, it will be for each individual afterwards to introduce as much of the disturbing affectionate element as he chooses, and to determine for himself the result on the new conditions supposed' ".

But Ruskin would have none of this. Instead of leaving the price of labour to the market place, and allowing workers to bid against each other, his solution is equal pay for any given job. If two workmen vie with each other to obtain a job, the employer should not exploit the situation to cut the rate, nor should he consider taking the less competent man for less money. Should any common scribbler, he was asked, be paid as much as a good author? And Ruskin's answer was "yes", if the man is given the job (adding that "the quantity of its money, which the country at present invests in scribbling is not, in the outcome of it, economically spent"), in other words the rate for the job should be fixed and then let the best man available get it. In more ways than one this is a fairly accurate description of current practice of wages administration at the shop floor level, with some superimposed adjustments to allow for skill and experience to be taken into account (except for unequal pay for women, which has persisted

Vol.
197

for generations for social reasons, but with the advent of equal pay for women many analysts have suggested that women will lose out in periods of unemployment as a consequence of the Ruskin doctrine). What astonishing foresight for Ruskin to have realized that a mechanistic view of treating labour merely as a commodity, left entirely to the mercy of market forces, is not going to work! He was, of course, severely criticized for failing to appreciate that a balance between supply and demand in a free market through the consideration of the marginal productivity of labour was the cardinal mechanism advocated by economists for securing full employment, and that any price of labour determined by means other than that balance would lead to unemployment. Ruskin makes several feeble attempts to answer his critics on this issue; that he did not succeed in providing a solution to unemployment is not surprising, since it continued to elude many generations of economists and politicians to come and has remained a live issue to this very day.

Nevertheless, Ruskin would not accept the notion of a floating rate for labour: "The false, unnatural, and destructive system is when the bad workman is allowed to offer his work at half-price, and either take the place of the good, or force him by his competition to work for an inadequate sum". For Ruskin this was a matter of social justice, not economics, and he went so far as to doubt whether the intrinsic value of any commodity—let alone labour—could be measured by its price. But these views do not lead him to propose equal pay for all: "if there be any one point insisted on throughout my works more frequently than another, that one point is the impossibility of Equality" and "my continual aim has been to show the eternal superiority of some men to others".

Vol. 4
1976

Thus, we face the eternal and so far unresolved problem: what is a job worth and how much should one pay for it? The adoption of equal pay for all does not, of course, provide an answer, while a system of pay differentials is bound to compound it. And since no way has been found to ascertain the intrinsic value of work, we have to conclude that the price can only be an outcome of a negotiating process which is not entirely immune from the effect of market forces. Even if labour is not to be regarded as a mere commodity, the ultimate output of labour is an array of goods and services, the price of which in a free market is subject to the relationship between supply and demand, to the implications of competition and to the degree of product substitution. Indirectly, therefore, the price of labour is inevitably affected, the pressure of the price for the final product acting as a constraint on the liquidity of the enterprise and on negotiations on pay.

It is for this reason alone that in our economic system equal pay for all is not a viable proposition. An even more persuasive argument is that equal pay for all ignores first the power struggle and secondly the role of incentives. The power struggle I have already alluded to: individuals or groups that have power are tempted to use it, they certainly want to preserve it, they often try to increase it. And while equal pay (with some adjustments) within a group is socially

acceptable, tactically advantageous and administratively convenient, the opposite arguments are easily advanced against parity between groups. As for incentives—these are vital in order to sustain both a corporate and an individual competitive spirit, a point frequently stressed by all groups of workers who value the acquisition of skill, the exercise of responsibility and the need for risk taking. What the warring factions in our society are not agreed upon is what these attributes are worth, hence what pay differentials are both just and necessary, and what consideration—if any—should be given to the needs of the individual, apart from the position that he would otherwise assume on an agreed spectrum value.

SAMUEL EILON
Chief Editor

REFERENCES

1. HUXLEY A (1929) Do what you will (see "*Revolutions*" *in Selected Essays*, Chatto & Windus, 1961),
2. RUSKIN J (first collected in 1862) Unto this last, four essays on the first principles of political economy (originally published in the *Cornhill Magazine* in 1860, later republished by George Allen, 1906).

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

FEEDBACK

The Myth of Self-Actualization—A Comment

I APPLAUD Professor Eilon for his editorial, '*The Myth of Self-Actualization*', (*Omega*, Vol. 3, No. 6, Dec. 1975) and agree with his observations, but would like to go a few steps further. Professor Eilon concludes (p. 630) "Self-actualization for all is a myth. It is doubtful whether it has any meaning as the ultimate goal for every individual, and whether such ultimate goals of spiritual well-being—even if they do exist—are identical for different people."

I wonder whether we should accept the concept of self-actualization (*sa*) in its present form. It has always bothered me because it is not paired with its other-self, which is self-control (*sc*). Further, *sa* is consistently treated as being something that is good (even though not attainable by everyone). What evidence do we have that the maximum potential sum of self-actualizations for any social group doesn't produce chaos? Do we know how individual *sa* achievements affect conflict? Could it be that anarchy is a likely, if not essential result of group *sa*? Can MAX Σ_i (sa_i), where i = people, produce a destructive result in the systems sense, and thereby, not only of the whole but of the people that are the parts?

[1] Presented at the 80th Annual Convention of the American Psychological Association, 7 September 1972, Honolulu, Hawaii; 'Consumenoid I—an explanation and an illustration' by M Lawrence Light; 'Applications of consumenoid I' by Kenneth A Longman; 'Saints and sinners—the meaning of the consumenoid model of human behavior' by William T Moran; 'Micro and macro aspects of behavior modelling: amplification of the entropy change notion underlying C_I's behavior' by Martin K Starr.

Management science is hardly in a position to include the present version of *sa* as a variable in its models. Yet solid theory development accompanied by empirical research could help clear the air about *sa*. At least, we could expand our understanding of this multiple-dimensioned objective whose name makes it sound as though it is one set of transitions along a singular path to the holy grail.

The management science approach allows the simultaneous possibility that Taylor's prescription for *sa* is a relevant compromise between organizational and individual objectives; and so is Herzberg's, and so is Fiedler's (who studies leadership). And so are the works of others (some completed, some in process, but most yet to be done). With respect to leadership, it does appear that many people are motivated to follow a cause; to join a team and/or to be a good soldier (not in the sense that Taylor spoke of soldiering). To accept a direct cause and effect relationship between work enrichment and *sa* overlooks other factors and ignores the characteristic procedure of management science to find a satisfactory or optimal set (in this case) of job designs. As another example of possible, interactive factors, there is evidence that change may be an essential component of what people seek, i.e. sa^1. Alternatively, what role does habit play in providing a fundamental sense of personal achievement?

Management, in the pursuit of socially beneficial productivity, cannot afford to accept the simple proposition that job enlargement and participatory management are the correct methods for securing *sa*, and thereby better industrial relations and improved productivity. It may just turn out

Vol. 4
1976

that in a relatively affluent society, *sa* without *sc* will mainly decrease productivity, even though such a result might at first be masked by the positive benefits of experimentation, *à la* the Hawthorne effect. (If there is an analogy between permissiveness in education and motivation to learn, then we already have evidence of the need to couple *sa* with *sc*.)

The desire to include human behavior in management science models had its first peak in the 1950's. Now, again, there is motivation to move in this direction. We are open to the possibility that there is no way to achieve ultimate productivity in a society that values the human qualities of all workers. If this is the case, then perhaps management should spend its time and creative efforts on learning how to: aggregate industrial and service systems that are large enough to warrant automation of all dehumanizing jobs; regulate such giant institutions so that they constantly work for the good of society; and provide the means for using leisure time to assist in the attainment of the many facets of *sa*. When the Taylor model is applied to the selection of the best machines, the past century of debate turns out to be preparation for the next century where work just might be the most pleasurable part of life.

MARTIN K STARR

Columbia University
Graduate School of Business
New York, NY
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Managers' Uses of Models

LEONARD S SIMON

CHARLES LAMAR

GEORGE H HAINES, JR

(Received September 1975; in revised form January 1976)

This paper argues that management scientists need to understand how models are used by managers and correspondingly apply appropriate criteria to their own efforts. Types of models used by managers are discussed. Then an example is presented of how such understanding can be used to evaluate the worth of formal analysis. The evidence points to patterns of quite different inquiry and response by managers than have typically been envisioned by management scientists.

Vol. 4
1976

TYPES OF MODELS

THE TYPE of model most often presented in the management science literature is the decision making model,[1] customarily defined as a logical or mathematical model used in a normative manner. We shall call this kind of model Type I. While the model may be designed to replicate the decision process of the decision maker [2], it is more common to construct a model which assumes: (a) a goal or set of goals, (b) logical expressions, often mathematical, which relate variables the decision maker can control, or forecast, to the goals, and (c) constraints on what it is possible to do. The problem is then to achieve the goals, given the relations (a) and (b). The common decision making model is so structured in terms of maximizing or minimizing the goals that they are attained by doing precisely what the model sets to do. This answer, a statement of what action will best achieve the assumed or given goals, is obtained by using a known computational technique sometimes implied in the model (e.g. linear programming). The literature in management science and operations research abounds with such models.

There is all too infrequently assessment of the Type I model. It is usually

[1] The term decision making is used in the same sense as Eilon [4 p. B173]: a multistage process commencing with 'information input' and terminating with 'resolution.'

an article of faith, at a 'practical' level, that the model will do better than a mere human. However, one early examination of this assumption showed that three different optimizing models for a common business decision problem, locating warehouses, were not able to produce decisions which were in fact optimal [11]. The models were not able to produce decisions as good as a simulation program which was not designed to necessarily produce optimal results. The reason for such a result is simple: the assumptions used in constructing the optimizing models were sufficiently incorrect compared to the structure of the real world to ensure non-optimal results. This problem apparently still exists [15].

A second type of model, labelled Type II, which is sometimes used is designed to aid a decision maker by supplying him with information which he would not otherwise have. Such a model is designed to convert raw data into a form that some humans can use as inputs into their decision processes. This type of model, which is occassionally discussed in the literature, may in fact be more frequently used than the Type I model. An illustration of this effect is that broad set of models which deals with the prediction of consumer behavior [12].

There may be several reasons for preferring and advocating a Type II model rather than a Type I model. First, it may not be possible to identify operational and meaningful goals compatible with formal analysis. Second, the model constructor may not feel he possesses adequate knowledge to build a reasonable Type I model. Third, from a strategic viewpoint a Type II model is easier to implement than a Type I model. Finally, in some situations there may be social mores against the use of a Type I model [1].

The foregoing are the types of models the meetings and journals largely address and discuss; it is probably not unfair to claim the emphasis is placed on Type I models. Work on models of these types is important and useful. The point is, however, that such models have only limited import relative to the needs of management of any large organization. Since the literature in management science largely ignores other potential valuable uses of models from a management perspective, the view has arisen in some quarters that management scientists are largely talking to themselves. This is probably correct if only the content of papers published in the journals and presented at meetings is considered, but probably incorrect if the reality of a successful working management scientist is considered.

Types I and II are most often the form in which the management scientist works. However, the manager who orders the formal analysis may have an entirely different conception of what the objectives of the analysis are. When the manager's objectives and the actual modelling work are coincidental in character, no problems arise and Types I and II exist in their pure form. When the objectives of the manager and the actual modelling work are not coincidental, three additional types of use of formal analysis arise.

The first of these other types of use is the development of a 'need to know' model, hereafter referred to as Type III. Such a model is designed to force, by

creating a need to know, one part of an organization which has data to give it to another part of the organization which wants it. To the manager, building a model establishing a need to know justifies from the corporate viewpoint acquisition of the data. The model has no intended use other than to create a communication link in an organization where none existed before, and is therefore most useful to a management scientist operating in a bureaucratic structure. Simulation models are often used for this purpose; one example of a Type III model of this kind is discussed by Davis [3].

A fourth kind of use, and one of great practical importance, is the model designed and used to provide a rationale for decisions previously reached by the manager based upon informal analysis. As Wilensky [14, p. 16] comments: "There is no doubt that much policy is improvised and later, with the help of experts, made to appear planned. Neither is there doubt that all experts . . . contribute to the self-esteem of their employers by lending their prestige and 'fronting' for them with various publics, 'taking the heat' when mistakes are made, and, in general, creating the verbal environment of the organization." The literature is replete with Type IV models presented as Type I or Type II—witness corporate planning models.

Vol. 4
1976

Type III and IV models are also important in the context of an organization's relation to its environment. They may, in the long run, lead to use of the more classical Type I and Type II modelling. For example, a Type III model may be required to get an organization to collect external data. Once the data is collected a Type II, or even a Type I model can be constructed to utilize the data. There is no evidence that the same model type can be successfully used to both ensure data acquisition and to process the data appropriately once it is acquired. A Type IV model might at first glance seem to be unworthy of serious consideration: after all, rationalization is not science. Unfortunately, reality is not so simple. Problems have a way of recurring. What was once a Type IV model may be converted into a Type I or Type II model as the problem recurs, as for example, has often happened with manpower planning models or plant location models. After all, such a model will have gained the confidence of the manager in its earlier Type IV application. The point is that the most common types of models, from a management perspective, may be the Type III and Type IV models.

A fifth type of model is one designed to establish the manager's reputation. In all probability, it will therefore look like another type acceptable as 'scientific' to a community deemed important by the manager. Unlike Types III and IV often all people involved in the construction of this model understand its purpose although it has not been explicitly stated. A Type V model is usually reported as an implemented application of Type I or Type II although this is rarely the case—the manager is unwilling to assume the inherent risks. The manager knows that corporate and personal goals are strongly intertwined. Although the discipline of management may suffer because of the development

of such models, in the eyes of many, this is a perfectly legitimate activity from a career objective viewpoint.

MANAGERIAL BEHAVIOUR PERTAINING TO MODEL USAGE AND EVALUATION

A taxonomy based upon the purpose for which management science analysis is requested as perceived and ordered by the manager was developed. Additional insights can be gained by further examining the managerial behaviour which surrounds the employment of each of the model types.

Why is knowledge of this behaviour so valuable? As Galbraith [8] and Vroom [13] point out, supervisory behaviour and participative decision making are the major tools at management's disposal in influencing the decision to produce.[2] This literature indicates that to achieve high quality useful output clear communication about what needs to be done and why is required. One way of effectuating such communication is for the management scientist to be able to understand what the manager wants not only in terms of the technical details, but also in terms of the implied model type as estimated by knowledge of managerial behaviour.

There are two points at which management interacts strongly with analytical work. The first is when the management inquiry is made to the management scientist. The second occurs when the management scientist presents a report to the manager of his work based on the original inquiry. At each of these points, behavioral interaction occurs between the manager and the management scientist. The theory [5, 8, 13] of what determines the decision to produce implies one must focus on significant differences in behavioral patterns to understand not just what is said, but also what is meant. These differences will help a management scientist to recognize the type of model which management desires. Table 1, therefore, presents the significant behavioral patterns associated with each of the five kinds of models delineated (or used) during these two interaction times.

Some brief examples of applications of specific model classes may help in tracing through the various behavioural patterns. Type I models are well represented by such techniques as mathematical programming to determine the fuel mix resulting from catalytic cracking in oil refineries. Type II are budgeting and control models which provide the decision maker with valuable information about performance. Many simulations, such as test models, pilot plants, etc., are also Type II models. Type III is exemplified by decision theory models used to justify collection of data. This is usually based on preposterior analysis

[2] The traditional thrust of normative application in the 'leader behaviour' literature [5] has been to attempt to modify supervisory behaviour. We are arguing that the same research can be used possibly to increase productivity by modifying 'subordinate' behaviour.

TABLE 1. TYPES OF MODELS EMPLOYED BY MANAGERS

	The character of the original management inquiry	The nature of the management response to the completed project
Type I	Decision-making Usually in the form of an operational problem requiring solution. Specific solutions are expected. An important goal is to replace current decision makers with lower level people working in conjunction with the model. An expectation exists that substantial underlying mathematics will be required to develop the model.	Review and acceptance of the work if the conclusions seem reasonable. Operational responsibility is delegated down the line for execution by subordinates. The model is expected to 'work' without significant restructuring for several years and with only very occasional review by management.
Type II	Information supplying The information which can be used in the personal decision making process is sought. The manager desires to operate in a 'what if' mode. Frequent use of the models is anticipated and rapid feedback required.	Analytical results used by present decision makers to supplement current procedures. Problem is broken into small parts so that computation and compilation can be delegated in the future.
Type III	Need to know The production of data, not necessarily in the form of numbers, for the manager's subordinates is the priority. Focus is on the current need rather than anticipated future problems.	A communication link is created where none previously existed. Data not previously collected or amassed, either internally or externally, is now done so on a regular basis.
Type IV	Rationale The goal is to develop a means (model) for justifying current managerial practices in a sophisticated, analytical fashion without criticism of any present policies or procedures. A desire for internal secrecy until the project is completed is expressed.	The 'concept' and its execution are presented to a public outside the organization. The work is often continued and expanded as the managers begin to lose sight of the original objectives and view the model in terms of Types I and II.
Type V	Reputation The types of inquiry associated with all four of the preceeding types of models are appropriate, but Types III and IV only if they may result in outwardly looking like Types I or II. The manager's motivation is usually presented in amorphous terms. Large scale spending is expected to give the project great significance.	Evaluation is primarily based on what others will think of the manager's capabilities rather than anticipated corporate benefits. Is the model developed in the forefront of available technology and will others take notice? Actual useful implementation is rarely contemplated. The analyst is encouraged to write a paper for the manager to present.

Vol. 4
1976

showing a high expected value of sample information. Type IV can humorously be described by managerial statements such as "The Penn-Central Railroad has substantial profit potential". Finally, Type V cannot be readily differentiated in terms of areas of application or apart from specific situations. It will always have the form of one of the other four classes. What does differentiate a Type V application is the specific behavioural character of the management inquiry and the nature of the management response.

There appear to be three general ways in which management attempts to evaluate the worth of management science models: (1) by management impressions of the value of particular work, (2) by examination of the rate of adoption of models elsewhere, and (3) through systematic costs–benefit analysis. All these evaluation techniques are sometimes useful. The point is, however, that they need to be placed in context. A careful justification for using formal analysis in a given situation is context dependent. The type of model use should be specified.

A SPECIFIC CASE

The problem[3]

The media selection problem is one of the most researched areas in advertising decision making, probably because of its highly quantifiable nature and its potential payoff. Much can be said about appropriate theoretical approaches to solving the media selection problem, but these ideas have never seemed to work out pragmatically as well as might be wished [10].

The research literature on this problem behaves in accordance with the typically expected Type I pattern. The problem is cast in a form that will permit use of Type I modelling despite the fact that the reformulation necessary to do this means the researcher is no longer solving a Type I managerial problem. The managerial problem is to maximize profits; the management science problem has become maximizing 'reach' or other proxi-variables. As a result, most models in this area that are used are used as Type II models.

The research reported below is an attempt to develop a media schedule for a small budget advertiser in the consumer realm. The question examined is whether media scheduling by formal analysis would be beneficial in this particular case, and what the context of use is in terms of the model typology presented above.

The company concerned is a savings bank of approximately 400,000,000 dollars in assets size. The product is deposits in savings accounts. The firm engages in two types of advertising: (1) general advertising to attract new customers or to get present customers to deposit more to their accounts, and

[3] A more extensive description of this case is available from the authors.

(2) special advertising to announce new branch openings, premium promotions and the like. This study focused only on the former.

The bank's advertising agency was considered by the bank's officers to provide excellent creative work. No past technical analysis for media scheduling had been done; the agency's officer in charge of media reported that he used the 'brewmaster' approach in his scheduling. It was hoped that by use of some established quantitative tools an improvement might be made in the area of media scheduling. Both the bank and the agency were eager to participate in the study. The bank's management had heard of studies of this kind and initially expected to be presented with formal analysis, probably of a Type I mien.

Analytical method

Based upon management's expectations, it was decided to utilize a mathematical programming approach, developing a nonlinear integer programming model.[4] The method chosen consisted of four steps:

1. definition of the bank's advertising objectives and the target audience group,
2. development of coefficients for the objective function,
3. determination of constraints, and
4. actual evaluation to determine the 'optimum' schedule.

The bank set as its goal obtaining large new savings accounts. Previous studies had indicated that savers who opened major accounts were generally over 45 years of age and/or had family incomes in excess of $15,000 per annum. A system of weights denoting the relative importance of income to age was then developed through discussions with bank officers to segment the target audience.

The three media most used by the bank, radio, television, and newspapers, accounting for 87% of the bank's media space costs, were all that were considered. A number of adjustments had to be made on the media data, such as on the age groups by which the available published data was broken down, to conform to the bank's target group and market area.

Only those types of insertions which were normally used by the bank in each of the three vehicles were considered—for example, 10 or 30 sec spot announcements and 5 min newscasts during drive time radio. Since no data were available on the relative effectiveness of insertions of various lengths on radio and TV, a subjective weight was used based upon the opinions of the vice-president in

[4] One possible alternative, that of a large scale simulation, was thought infeasible since the bank did not have the resources necessary to develop the data for such a simulation.

Vol. 4
1976

charge of media scheduling and the account executive at the agency. Cost data were taken from the Standard Rate and Data Services Manual and adjusted where necessary, to conform to the average audience estimates.

The foregoing estimates generally lacked information on a variable of major importance, target group income. This was estimated by adjusting *US Census* data producing a figure for the estimated target audience, but it contained the major problem of not differentiating adequately among individual media.

Finally, the relative impact of each type of media for any product and the relative climate for the bank's purposes were assigned subjective weights based on an average of the estimates of the bank's advertising manager, the marketing vice-president of the bank, and the agency account executive.

It had been necessary up to this point to bring management into the development of the model at several stages where subjective evaluations were required. Management willingly supplied this information but began to question how the output of the model could be definitive given all the judgmental inputs. The researchers concurred with management's feelings, and the goal of the formal analysis shifted away from a Type I effort towards a general evaluation of the agency's media scheduling (Type II use) and also towards an evaluation of collecting special information outside the bank to solve some of the data problems discussed above (possibly a Type III use).

The objective function and the appropriate constraints took the form:

$$\text{Max} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{WEE}_{ij} \left(1 - e^{-ax_i}\right)$$

$$\begin{aligned} \text{Subject to} \quad & \Sigma_i c_i (x_i) j \leq B \\ & x_i \leq L_i \qquad\qquad i = 1,\ldots, n \\ & x_i \text{ integer and} \geq 0 \qquad i = 1,\ldots, n \end{aligned}$$

where:

- WEE_{ij} = weighted effective-exposure coefficient of insertion j in vehicle i,
- x_i = number of insertions in vehicle i,
- a = parameter describing curve shape; reflects diminishing effectiveness cumulatively over all insertions in one vehicle,
- $c_i(x_i)$ = cost of x_i insertions in vehicle i,
- L_i = upper bound constraint on number of insertions in i^{th} vehicle (e.g. 52 issues of the Sunday paper),
- B = advertising budget.

Actual rates were used; thus the cost structure is nonlinear with abrupt break points. The nonlinear integer nature of the model renders infeasible any simple algorithmic solution. Therefore, direct search was used to maximize

weighted effectiveness exposures. The procedure is defined by the GROPE program [6, 7].[5]

Table 2 presents the basic results. Column 1 evaluates the agency's proposed budget; column 2 shows that using the model and the same budget, 310,000, additional WEE units can be achieved.

TABLE 2. MEDIA SCHEDULES

Vehicle	(1) Agency results $70,200	(2) Model results $70,200	(3) Model results budget of $77,850	(4) Model results budget of $82,500	(5) Model results budget of $54,650
1	28	56	137	154	0
2	156	81	0	21	23
3	104	3	267	259	154
4	312	53	1	21	0
5	21	0	0	0	3
6	19	0	0	0	0
7	19	21	23	30	0
8	80	160	122	129	180
9	80	46	123	118	126
10	6	43	25	24	15
Budget Used	$70,035	$70,200	$77,846	$81,805	$54,647
Weighted effective exposures WEE(000)	2585	2895	3306	3393	2612
$ Cost/1000 WEE	$27·09	$24·25	$23·54	$24·10	$20·92

Table entries are number of insertions.

Now the question is: what is the value of this analysis? There are two possible ways to approach this question. The agency had originally proposed a budget of $11,800 for creative purposes. If the bank were to use the same ads for another year, this money could be spent on additional insertions. Column 4 shows that if this were to be done, a gain of essentially 500,000 WEE units would be achieved. However, the nonlinear nature of the problem is indicated by the result in Column 3, which achieves a gain of 410,000 WEE units and still leaves $4150 for creative work. The second possible way to place a value on the additional 300,000 WEE units is to ask what budget, using the model, would achieve the agency's WEE figure. Column 5 indicates that a budget of $54,650 would essentially provide the same WEE units. Thus, if the model were

[5] GROPE has four different direct search subroutines and a control subroutine that chooses one of these four, based on past experience. A linear stochastic learning model is used to provide the adaptive choice mechanism. The four direct search methods used are: Hook–Jeeves pattern search, simple random search, shrinkage random search, and Satterthwaite search. Such a procedure, of course, does not necessarily identify a global maximum. Its chances for doing so depend largely on careful use of the direct search program.

Vol. 4
1976

to be used in a Type I fashion to achieve the agency's implied WEE goal, the model's implicit value would be about $15,500.

There is another way in which the costs of achieving a given level of WEE could be further reduced. There are set up costs associated with using any given medium in terms of production costs and working out scheduling arrangements. Where only a very small number of insertions is to be scheduled, the medium could be eliminated. However, given the nature of integer programming, it is not at all certain that the optimal schedule would remain even approximately the same with one or more media removed. Thus, this is another area in which managerial judgment would be quite important and in which the model would be most helpful in determining the results expected from given courses of action.

The costs of developing the model including salaries and information collection were about $6000. The total cost of the effort was not large and compared favorably with the imputed cost savings from using the model (a Type I evaluation). In addition, since in future years the only costs would be those of operating the model, collecting data, and making any improvements desired, the benefits from the model should be even greater.

Evaluation

It is relatively rare that management scientists attempt to apply economic criteria to the question of whether they should have worked upon a project. When they do so, they are very likely to assume that their goal was to construct a Type I model. In the context of the above discussion, such a goal would imply removing media scheduling decisions from the set of tasks performed by the agency. Since the agency would still perform all the other tasks associated with the general problem, the measurement of improvement in performace over the agency's previous 'unscientific' approach is biased downward. Thus, from management's viewpoint, the problem could never be viewed as amenable solely to a Type I model solution.

A better understanding of the ways in which management science work is viewed by managers would lead to the evaluation of the example described in terms of Type II and Type III models. Management rarely expects formal analysis to be justified on the basis of replacement of human beings' decision-making competence. A study such as this provides management with an opportunity to re-evaluate its confidence in the agency and its contributions. Thus, management could operate under a Type II model with regard to a different objective, namely the quality of the agency's work. In this case, the model's economic contribution might have been identified as equivalent to the cost of having an outsider review the agency's work. Hiring a consultant who would write a report evaluating the agency would not, however, have given management an improved understanding of the structure of the problem, the additional information required to use 'sophisticated' analytical models, and the ability to evaluate what-if questions, as does the model.

Alternatively, the work herein could have been seen as an effort to evaluate and/or force collection of data, a Type III model. Essentially, what the analysis did was highlight areas where better information was needed regardless of whether a formal analytical approach was utilized in the future or decision-making continued on a judgmental basis.

IMPLICATIONS FOR THE MANAGEMENT SCIENTIST

Failure to understand how managers use models can lead to over simplified economic analysis of the benefits of a given model. The fact that managers frequently do not agree with researchers' conclusions about the benefits of a model does not mean the managers are stupid or naive; they are just marching to a different drummer. The point is that cost and benefit figures or merit can and should be produced but to decide whether a model is justified in a particular context remains a matter of management judgment and interpretation.

How does the management scientist operate successfully in this environment? First, he must accept that the goal of his research is not necessarily finding the 'right' answer for that may not be what management wants, e.g. Type III, IV, V. Second, he has to understand the broad role of management science in the organization.[6] Management scientists would do well to consider carefully what the expectations and goals of management are in supporting research and not blithely assume all efforts should be evaluated in terms of attaining a blissful state of optimization.

Vol. 4
1976

ACKNOWLEDGEMENTS

The authors are indebted to Charles Meirsch for many helpful suggestions and discussions and to Martin Evans, Gerald Stiles, Tsvi Ophir and Venkataraman Srinivasan for helpful comments.

REFERENCES

1. BURDICK E (1965) *The* 480. Dell Publishing Co., New York.
2. CLARKSON G PE (1962) *Portfolio Selection: A Simulation of Trust Investment*. Prentice-Hall, Englewood Cliffs.
3. DAVIS RK (1974) Interactive gaming: a production example, in *Simulation Games and Experimental Learning Techniques* (Eds. KENERDINE J & KEYS B). Center for Economic and Management Research, Norman, Oklahoma.
4. EILON S (1969) What is a decision? *Mgmt Sci. Applications*. **16**(4), 172–189.

[6] For some insight on this subject see the comments of Michael Radnor and the other panelists in [9].

5. EVANS MG (1970) The effects of supervisory behaviour on the path–goal relationship. *Organizational Behaviour and Human Performance* **5**, 277–298.
6. FLOOD MM & LEON A (1964) *A Universal Adaptive Code for Optimization (GROPE)*. Working Paper No. 18, Space Sciences Laboratory, University of California.
7. FLOOD MM & LEON A (1966) A universal adaptive code for optimization (GROPE), in *Recent Advances in Optimization Techniques* (Eds. LAVI A & BOGL TP). John Wiley, New York.
8. GALBRAITH J (1967) Influencing the decision to produce. *Ind. Mgmt Rev.* **9**(1).
9. HALBRECHT H, SAVAS ES, HOFFMAN G, AYRES HF, RADNOR M & EDELMAN F (1972) Through a glass darkly. *Interfaces* **2**(4).
10. KAPLAN RS & SHOCKER AD (1971) Discount effects on media plans. *J. Advertising Res.* **11**(3), 37–42.
11. KUEHN AA & HAMBURGER MJ (1963) A Hueristic program for locating warehouses. *Mgmt Sci.* **9**(4), 643–666.
12. MASSY W, MORRISON DG & MONTGOMERY DW (1970) *Stochastic Models of Buying Behaviour*. MIT Press, Cambridge, Massachusetts.
13. VROOM VH (1960) *Some Personality Determinants of the Effects of Participation*. Prentice-Hall, Englewood Cliffs.
14. WILENSKY HL (1967) *Organizational Intelligence*. Basic Books, New York.
15. WOOLSEY RED (1972) A candle to Saint Jude, or four real world applications of integer programming. *Interfaces* **2**(2), 20–27.

ADDRESS FOR CORRESPONDENCE: *Professor LS Simon, Graduate School of Management, University of Rochester, Rochester, NY 14627, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Scheduled and Charter Impairment in North Atlantic Air Services

WILLIAM G BROWNE

(*Received August 1975; in revised form January 1976*)

The intent of the paper is to: (1) review the material leading to President Nixon's request that the Civil Aeronautics Board prepare economic criteria for judging degrees of impairment that exist in either charter or scheduled air services: (2) construct a model for interpreting the economic relationships between scheduled and charter services; (3) provide evidence that impairment currently exists and is causing market distortions; and (4) indicate that recent board decisions provide corrective influences on the distorted conditions in the passenger market.

Vol. 4
1976

INTRODUCTION

THE CIVIL Aeronautics Board issued on 20 April 1972 its opinion on the Transatlantic Supplemental Charter Authority Renewal Case (Order 72-5-9). Correspondence between the Executive Office and the Board, after the Board's original decision, (24 January, 1972), but before the final order, centred on the decision's compatibility with President Nixon's June 1970 "Statement of International Air Transportation Policy of the United States." The focal point was the Board's interpretation of the concept of 'substantial impairment.' President Nixon indicated that there should be more effort to establish economic criteria for judging questions of impairment in future cases.

The "Statement of International Air Transportation Policy of the United States" contained a section on the role of the supplementals' charter operations in relation to scheduled services, and vice versa.[1] 'Substantial impairment' was a term originated and used to reflect the economic effects that regulations concerning one of the service types would have on the other. Although criteria for examining or characterizing these effects were not defined, the statement is

[1] "Statement of International Air Transportation Policy of the United States," (The White House, 22 June, 1970) pp. 5–7.

important as the most current enunciation of United States international air transportation policy.

The term 'substantial impairment' probably evolved from the Transatlantic Charter Case of 1964 in which the relationship between the service types was qualified by a measure called 'substantial damage.' The concern in that case was the diversionary effect that supplements and/or charters would have on the scheduled services. Substantial damage, in terms of passenger diversion, became the measure of a new service's impact. The Executive Office and the responsible agencies appear to desire a broader concept to reflect a more complete relationship than merely passenger diversion between the service types. Thus the birth of 'substantial impairment.'

The intent of this paper is to analyze the background and issues that may influence or be influenced by any interpretation of the term 'substantial impairment' and a model for viewing the relationships between the scheduled and charter services in terms of impairment is developed. Using material developed in the study, the author will attempt to show a possible division of the services between charter and scheduled that will simultaneously provide the traveler with adequate alternatives of service and fares.

HISTORY OF AIRLINE GROUPS AND THEIR INTERPRETATION OF SUBSTANTIAL IMPAIRMENT

Pan American Airways (PAA) was issued its certificate of public convenience and necessity (PC&N) for transatlantic air service in 1939. In 1940 American Export Airlines was also given a certificate based on the rationale that Pan American Airway's monopoly position was not totally in the public interest. In 1945 Transcontinental and Western Air (TWA) was also given a certificate and all certificates for the North Atlantic were limited to a term of seven years. By 1950 PAA had acquired the North Atlantic PC&N certificate and properties of American Overseas Airlines, formerly American Export, which was primarily owned by American Airlines. Both TWA and PAA's certificates were then amended to provide for competition at major European traffic points.

Until 1955, TWA, PAA and the certificated foreign-flag route carriers carried practically all of the North Atlantic traffic on either their scheduled services or their charters. The right of first refusal for charter service by certified route carriers was eliminated during 1955 and the irregular carriers were permitted to provide charter services for legitimately derived groups at fair and reasonable rates. The supplementals, however, still had to file with and obtain prior approval from the Board for each international charter.

During 1964 two of the supplemental carriers were given certificates of PC&N for the North Atlantic and the supplementals' prior approval stipulation was discontinued. They were also given the rights to operate split charters

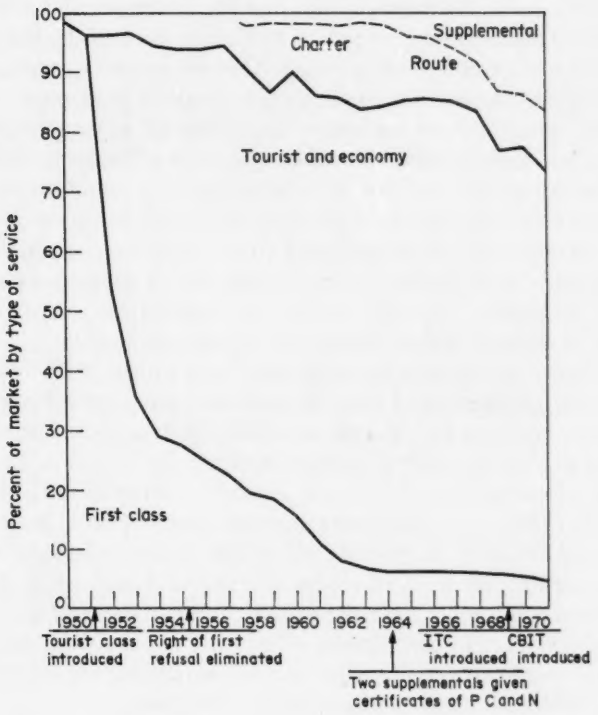


FIG. 1. *Percentage market shares by various types of North Atlantic air carriers, 1950–1971.*

so that separate charter groups could contract for portions of a charter aircraft. It was during the Transatlantic Charter Case hearings (Docket 20569 and Order E-24241, 30 April 1964) that the concept of substantial damage was introduced as a criterion for viewing the impact of added services on the North Atlantic system. In 1966 four more supplementals were given certificates of PC&N and permitted to operate transatlantic charters. At this same time, the supplementals were given exclusive rights to provide Inclusive Tour Charters. The growth of the supplementals exceeded the industry average starting with the 1964–66 period and the supplementals as a group gained a larger share of the total market (See Fig. 1).

In late 1969 the Executive Office organized a steering committee to formulate an up-to-date policy to guide the responsible government agencies in their decision-making in matters related to international air traffic. Many market conditions, both from the supply and demand sides, had changed substantially

Vol. 4
1976

since the 1963 policy statement.[2] The changes leading to the need for new analysis were typified by the tremendous traffic expansion of the North Atlantic supplemental airlines and the rising demand for bulk or mass transportation.

The 1970 policy statement used the term 'substantial impairment' to portray the negative impact that rulings concerning scheduled services would have on charter services, and vice versa. However, symptons of impairment for the two groups of service are viewed differently. The effects of charter rulings on scheduled services are viewed in terms of the ruling's impairment of regular and dependable frequent schedules, maintenance of routes including low density routes, traveler flexibility in departure and return times and/or flexibility in duration of the stay. If the charter operators were to obtain an inequitable market share, the scheduled carriers' ability to provide the schedule regularity and origin–destination opportunities may be substantially diminished. Numerous travelers would then be inconvenienced and may have to modify their travel schedules to meet charter requirements. The effects of scheduled rulings on charter services evaluated in terms of the ruling's impact on the accessibility of low cost service provided by plane load movement, and a group's ability to tailor the services to its unique origin, destination and stopover needs. This indicates that the charter carriers' capacity to provide plane load prices and to fill varied off-route origin–destination needs of groups of travelers should not be diminished. Groups of travelers should have the ability to obtain quantity discounts and bargain for services to fit their desired schedule. The policy also indicates that both the scheduled carriers and supplemental carriers should be permitted to compete for the bulk (i.e. charter) transportation business.

FURTHER STUDY OF THE ISSUES NEEDED

Historically, impairment has been measured by estimates or traffic diversion. Using this measure, the passenger losses of one type of service correspond to the gains of the other type. The argument presented for or against proposed additions or deletions of service assumed a fixed group of potential passengers that were to be divided among various service categories.

A degree of sophistication was added to the diversion phenomenon by the introduction of the concept of price elasticity, supported by data showing the effect of changes in price level on the number of potential passengers. However, criteria for judging a service proposal's feasibility prior to the 1970 statement still focused mainly on diversion with only partial recognition of traffic genera-

[2] The 24 April, 1963 "Statement on International Air Transport Policy" was an update of the policies developed in 1944. During the nineteen years between policy statements, substantial political, equipment and traveler changes had taken place in the international air transport industry. The modified policy statement was to reflect the needs generated by these changes and to provide concise direction to the government officials in carrying out their statutory responsibilities.

tion caused by price decreases. The 1970 statement, however, brought a need for improved models able to take into account simultaneously overall market size, service quality, and carrier welfare. Many symptons of a distorted or even demoralized North Atlantic market appeared in the early 1970's, such as strandings of charter passengers, illegal ticket discounting, illegal charter formations, bankruptcy of some of the supplemental carriers and travel agencies, instability of the IATA fare structure, and distortions of bilateral agreements. Underlying these conditions were: an oversupply of available aircraft, the addition of new foreign charter carriers, improved consumer knowledge, the public's demand for lower prices, and the commitments of governments (particularly the U.S.) to provide the public with a variety of service and price packages.

Regulators will be responding to the problems listed above and attempting to satisfy the spirit of President Nixon's policy in several current and impending cases. To provide consistent action throughout the individual hearings on each of the cases, the regulators will need to provide a useful interpretation of the term *substantial impairment*.

Vol. 4
1976

DEFINITION OF IMPAIRMENT

Impairment is defined in terms of dollars of revenue not realized by a service type (charter *vs.* scheduled) as a consequence of a regulation, ruling or Congressional act that restrains that service type from obtaining an optimal market share. Dollars of revenue not realized are the difference between actual revenue and the potential revenue under monopoly conditions with only one type of service. Potential revenue is the revenue that could be obtained by one of the service types provided by any number of carriers in an unregulated market containing only that one category of service.[3]

The same approach can be used to find revenue potential of a system consisting entirely of charter services. Charter revenue forgone is determined by substracting the charter revenue that could be obtained under a mixed charter and scheduled system from potential revenue for charter service alone.

[3] For example, if only scheduled services were permitted in the North Atlantic market between the United States and Europe, the potential revenue of this system for 1971 can be estimated. During the Supplemental Renewal Case (Transatlantic Supplemental Charter Authority Renewal Case, Order 72-5-9,) it was estimated that about 1/3 of the charter passengers would have traveled by scheduled service if charter services had not been available. Accordingly, 6.7 million scheduled passengers in 1971 plus 1/3 of the 2.5 million charter passengers would have produced 7.6 million passengers. At an average one-way scheduled fare of $205 potential revenue would have been $1.6 billion. Thus, revenue forgone or dollars of revenue not realized for scheduled services is the differences between estimates of maximum potential revenue of $1.6 billion and what could have been obtained under other regulatory conditions.

OVERVIEW OF MODEL

To attempt to capture simultaneously the influence of many of the impairment interpretations and view the total impact on the size and division of the North Atlantic market three graphical models are developed, two of which relate the number of passengers using a service type, scheduled or charter, with a level of income forgone, while the third model relates scheduled traffic to charter traffic. A fourth model indicating a relationship between the revenue forgone between the two service types is then developed. This will provide an algorithm for tracing the economic consequences of rule changes concerning one service type on the other service type. For instance, the analyst can gain more insight into the possible increase in charter demand if family fares are discontinued on scheduled flights.

A point of zero impairment for both service types is established in the final model and deviations from this point are discussed. The concept of 'substantial' is developed using historical growth data and is used to show that substantial impairment does exist. Although the data used in developing the model are limited to the North Atlantic market (the largest charter market) between the United States and Europe, the concepts should be widely applicable.

Scheduled revenue forgone

Revenue forgone of the North Atlantic Scheduled services is related to the number of passengers who travel by scheduled services in Fig. 2 (where scheduled revenue forgone is the potential revenue not realized due to some regulatory

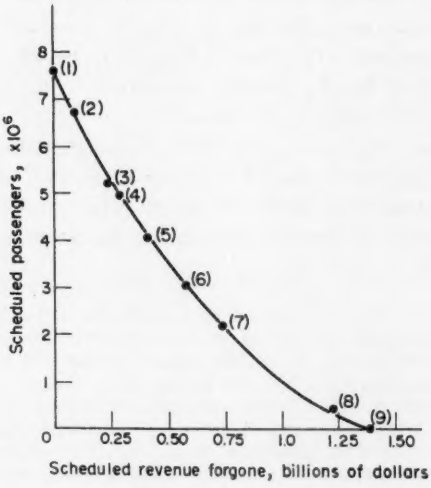


FIG. 2. *Scheduled passengers vs. scheduled revenue forgone for the North Atlantic market.*

condition) shows that relationship between scheduled revenue forgone and the number of passengers that would use scheduled services given certain regulatory constraints. The data pertain to the North Atlantic market between the United States and Europe.

Using the lowest scheduled fare, the revenue forgone because of the competition developed by the charter services offered in 1971 yields the scheduled revenue forgone that corresponds to the 1971 traffic division. The scheduled traffic on the North Atlantic void of competition from charters would have increased by 837 thousand passengers, resulting in revenue forgone of $88 million.[4] Developing further revenue data requires that various segments of the scheduled traffic be deleted. Possibly the simplest method, the one used here, is to delete the various discount fare categories with the largest discount fares being eliminated first. It is assumed that the traffic that used these scheduled fares would then find it more convenient to travel on charter services—especially with the expanded charter services that would appear as a consequence of the improved market conditions caused by scheduled service cutbacks. The largest discount fares are usually used by individuals who have greatest flexibility in their travel schedules, thus the least need for a scheduled system.

Fig. 2 was produced by fitting a line to nine points taken from columns (f) and (g) of Table 1: Point (1) represents the situation of no restraints for the scheduled services.[3] Point (2) represents the existing 1971 market condition with the scheduled services losing approximately $8788.9 million in potential revenues.[4] Point (3) represents the 1971 condition less the family plan, military, youth and/or other standby passengers (the largest of the discounted fares) that generated approximately $161 million in scheduled revenue during 1971. Point (4) represents the situation specified by (3) with the Affinity Tour groups eliminated. Point (5) eliminates the remaining tour group passengers, bringing the scheduled traffic down to 4 million and increasing forgone revenues to $417 million. Points (6) and (7) provide for the elimination of the 17–28 day excursion fare and the 29–45 day excursion fare respectively. Added forgone revenue would then come from the elimination of the full fare passengers, first by the economy fares represented by point (8) and then by the first class fares represented by point (9).

Charter revenue forgone

There are a number of methods that can be used to identify the relationship between charter traffic and charter revenues forgone. The method presented here (the solid line in Fig. 3) is probably among the most conservative. It will serve as an extreme indicating the minimum required change between service types. The other methods (the dashed line being an example) complement the conservative model in indicating the direction of needed change but requiring

[4] Using the lowest peak one-way fare of $105 the revenue forgone is estimated to be 837,333 × $105 = $88 million.

TABLE 1. NORTH ATLANTIC SCHEDULED REVENUE FORGONE VERSUS SCHEDULED TRAFFIC BASED ON 1971 DATA

	Scheduled market					
	(a) Passenger breakdown (per cent)	(c) Traffic by fare class (a) × (b) ($\times 10^3$)	(d) One-way fares	(e) Revenue by fare type (c) × (d) ($\times 10^6$)	(f) Accumulated revenue forgone ($\times 10^6$)	(g) Remaining scheduled passengers ($\times 10^3$)
(1) Monopoly	not applicable				0	7,573
(2) Charter	not applicable			$88	$88	6,736
(3) Other	22.8	1,536	$105	$161	$249	5,200
(4) Affinity tour	4.4	296	$138	$41	$290	4,904
(5) GIT/CBIT	12.5	842	$151	$127	$417	4,062
(6) Excursion, long	15.1	1,017	$156	$159	$576	3,045
(7) Excursion, short	12.8	862	$192	$165	$741	2,183
(8) Economy	26.0	1,751	$276	$483	$1,224	432
(9) First	6.4	431	$391	$169	$1,394	0

(b) Scheduled traffic between the United States and Europe for 1971 was 6,736.000 passengers.

Notes

Data taken from Docket 20781 and Order 72-2-104. The fares represent the peak season fares between New York and London. For (1) see footnote 3. For (2) see footnote 4. Scheduled revenue forgone is equated with the accumulated income forgone by eliminating each fare in the listed sequence. (3) covers youth, family plan and military fares.

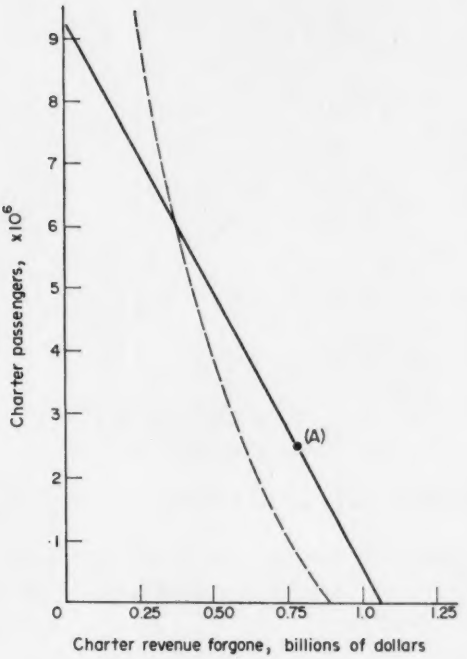


FIG. 3. *Charter passengers vs. charter revenue forgone.*

a much higher magnitude of change to obtain an optimal market division between service types.

Two points were used to locate and identify the charter revenue forgone and charter passenger line. First it was assumed that at zero revenue forgone nearly all of the passengers traveling the North Atlantic in 1971 would have taken charters. At zero charter revenue forgone the charter rules would have been liberalized to make charter formation easier and scheduled flights would not have been allowed to fly North Atlantic routes. Given a market served entirely by charter carriers, the 6.7 million scheduled passengers and the 2.5 million charter passengers would have flown on charter flights. The total number of passengers that would have flown on charters, given zero revenue forgone, is, therefore, approximately 9.2 million (see Fig. 3). Some of the passengers who originally did travel on scheduled flights certainly would have not gone if only charter were available, but most of these passengers would have been replaced by travelers who would take advantage of the wider availability of charters and/or lower average prices.

The average 1971 North Atlantic charter round trip fare was approximately $230 or $115 one way, corresponding to a maximum charter forgone revenue

Vol. 4
1976

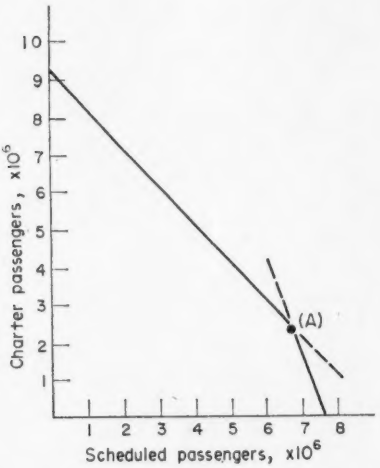


FIG. 4. *Charter passengers vs. scheduled passengers in the North Atlantic market.*

at zero charter traffic of $1.06 billion (the lower right hand point on Fig. 3). The 1971 charter revenue forgone is approximately $775 million ($115 × 6,736,000), shown as point A.

Scheduled versus charter traffic

A passenger trade-off relationship was constructed by identifying the actual 1971 traffic breakdown (point (A) in Fig. 4) and then making assumptions about the switching phenomenon as capacity availability moves from that point. The 1971 scheduled traffic between the United States and Europe was 6.7 million and the charter traffic was 2.5 million. In the Transatlantic Supplemental Charter Authority Renewal Case[5] it was concluded that 30 to 35 per cent of the North Atlantic Charter passengers would have traveled by scheduled service if charters had not been available. Using a percentage of −33.3 the slope of a trade-off line would be −3. Movement leftward from point (A) is characterized by a one-to-one trade-off line with a slope of −1. This slope signifies that for passenger deducted from scheduled services there is a passenger added to charter services. For example, if the youth stand-by fares were cancelled and charter regulations were liberalized to further accommodate this market segment most of the stand-by traffic would switch to the charters.

Revenue forgone and traffic switching between service types

Using Fig. 4 as the link, Figs. 2 and 3 can be related to one another. If the axes in Fig. 2 are reversed, the changes in forgone revenues of one service type can

[5] *Transatlantic Supplemental Charter Authority Renewal Case* (Docket 20569) Recommended Decision of William J Madden, Hearing Examiner, released 12/21/70, p. 2.

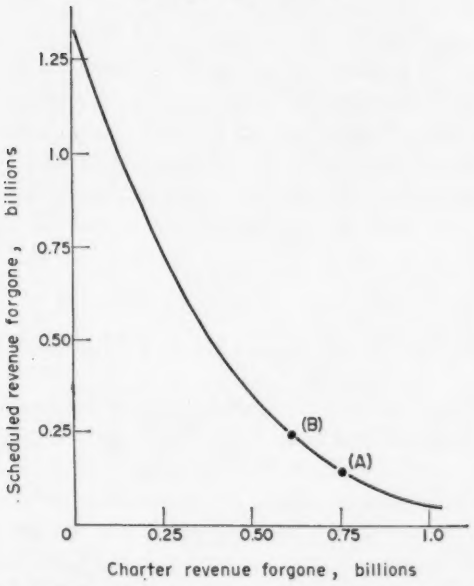


FIG. 5. *North Atlantic revenue forgone—scheduled vs. charter.*

be reflected into the other service type by using Fig. 4 as a 'mirror.' Thus added forgone revenues and traffic losses of one service can be transformed to traffic gains and reduced forgone revenues of the other service. Thus, a graph of scheduled versus charter revenue forgone can be developed from Figs 2 and 3 as shown in Fig. 5. Since this curve was developed from data reflecting the public's demand responses to different mixes of services, it can be regarded as a passenger indifference curve.

By reflecting point (A) in Fig. 4 representing the 1971 charter traffic of 2.5 million and scheduled traffic of 6.7 million through Figs. 2 and 3 to Fig. 5 a corresponding charter revenue forgone of $775 million and scheduled revenue forgone of $76 million can be found. Point (A) on Fig. 5 corresponds to the 1971 forgone revenues of charter and scheduled service. Together the two services generated traffic of 9.2 million and total forgone revenues of $851 million. If the operating point was moved to the left by a small increment total forgone revenue would drop with total traffic remaining constant.

There is a limit to the left hand movement since an optimal point will be reached and total revenue forgone would then increase. This optimal point (the point of minimum system revenue forgone), can be found by sliding a line at $-45°$ slope from the upper right hand side of Fig. 5 toward the origin, the point of last contact with the indifference curve being point (B), being the point of minimum system forgone revenue and should provide data for identifying an optimal division of the North Atlantic market between the scheduled and charter services.

From the revenue forgone indifference curve it can be seen that moving to the optimal point would involve decreasing the forgone revenue of the charter carriers by $139 million and increasing the forgone revenue of scheduled carriers by $114 million. This would yield an increase of one million passengers in charter traffic and a decrease of one million in scheduled traffic. Thus the net effect of the change would be to lower system forgone revenue by $26 million to $825 million and shift 1.0 million passengers from scheduled to charter services.

CRITERIA OF SUBSTANTIAL IMPAIRMENT

The terms 'impairment' and 'substantial' can now be introduced and used to evaluate the meaning of deviations from the optimal market division. Since revenue forgone is going to exist for both service types as long as they are permitted to coexist, it is necessary to select a base point and to measure impairment as deviations from the basing point. The optimal market division, in terms of revenue forgone, is identified as the point of zero system impairment and zero impairment for each service type. Fig. 6 is a modified version of Fig. 5, indicating the consequences of shifting the axis to the point of optimal market division. Deviations from the base point are measured in billions of dollars

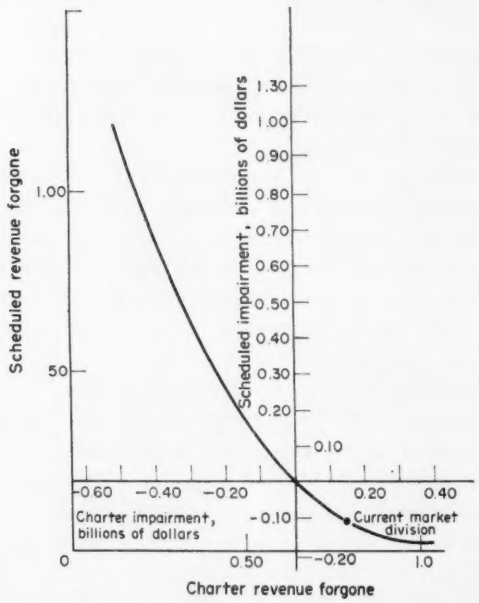


FIG. 6. *Charter impairment vs. scheduled impairment for the North Atlantic market.*

and called impairment dollars. Point (B) in Fig. 5 becomes the origin for Fig. 6 and the point from which deviations can be measured. With the axes change it will be easier to visualize the impairment changes that will have to be undertaken to move towards the optimal market division from the current division.

'Substantial' will be defined as any point of division beyond those that could occur, with 95 per cent confidence, under natural market conditions. The standard error of the traffic growth was used to indicate the range of traffic division that could exist and be accounted for, with 95 per cent confidence, because of random market behavior. The standard error for scheduled market growth is 1,126 passengers. The 95 per cent confidence interval or two standard errors would require that the current market division be within $\pm$2,252 passengers of the optimal division in order to assume that their difference was insignificant. As noted earlier the existing deviation of actual from optimal division is 1.0 million passengers, indicating a substantial impairment. Another way of providing the same conclusion is to transform $\pm$2,252 passengers to dollars of impairment via the equation for the curve in Fig. 2. This transformation indicates that substantial impairment exists if the difference between actual and optimal charter or scheduled impairment exceeds $690,000. Since the differences were in the range of $100 million it is easy to identify the existence and approximate magnitude of substantial impairment.

Vol. 4
1976

It can be concluded that in 1971 the charter services were impaired substantially. To correct for this impairment would require two classes of action. First the rules governing the charter carriers should be liberalized to lower the charter impairment. Since 1971 substantial rule making has evolved which benefits the charter carriers. This, of course, would be in the form of the new TGC service which was introduced in 1973. Secondly, it is necessary to move the market division toward the optimal point concerns the scheduled services. The non-profit generating fares, which may be providing services for a group of travelers who have a minimum need for the scheduled services, should be changed. The form of this change should be the discontinuation of the extreme discounting for specific market segments (youth, family, military) that have little real need for scheduled services. The airline expected response from the scheduled carriers would be to reduce scheduled services to generate a load factor that reflects the demand of the higher fares and not the current demand that includes passengers not requiring scheduled services.

OPPORTUNITY COST OF SUBOPTIMAL OPERATION

The model presents a number of items that appear to be consistent with most intuitive thought: (1) The extreme position of having only scheduled service is a substantial cost for the public and for the charter carriers; (2) the current position is not one that best serves the public interest and the public's capacity

to afford travel on the North Atlantic; (3) regulation that reduces the suppression of the charter growth will cause market shares to move closer to an optimal division; and (4) providing all charter services would not be a solution to an optimal division and would not yield a system of charter and scheduled services that would best reflect the integration of public interest.

It appears that the Board is continuing to liberalize the regulation governing the charter services.[6] The process of methodically reducing the barriers to charter traffic over an extended time period will allow the system division to move towards a more optimal position without substantially disrupting the participants. It will place pressures, however, on the airlines to modify their services to provide the service types demanded under the regulations. If the necessary changes are not made continuously and in an anticipatory fashion the penalty will be in the form of losses in market share and lower profits. Without continuous changes an airline may have to make drastic changes to recover parts of their market share and profits that have eroded.

Regulation can be adjusted to stimulate the shift to a system providing substantial freedom for introduction of charter services. Since all the regulations will not change at once, but over an extended period, the scheduled airlines will not have to make an extreme and unmanageable shift in operations. They will have to use some foresight, however, and make the appropriate shifts as the public awareness of charter alternatives grows and charter demand swells.

The evidence presented indicates that substantial gains could be made from the standpoint of the entire system if the participating airlines and regulators would provide the capability and incentive to move closer to an optimal market division.

ADDRESS FOR CORRESPONDENCE: *William G Browne, School of Business, Oregon State University, Corvallis, OR 97331, USA.*

[6] The September 1972 ruling to introduce Travel Groups Charters is an example of action that should provide incentive for further charter expansion.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Linear Programming and Financial Cost Models for Industrial Enterprises

MICHEL ARATEN

BERNARD DICKMAN

(*Received August 1975; in revised form December 1975*)

Linear programming models of complex interrelated manufacturing enterprises have improved production planning decisions. With a linear programming formulation, the planning engineer concerned with specifying levels of production activity need not be concerned with cost accounting formulations of product costs to make his production decisions. He is only concerned with examining alternative production activities in terms of the added production costs the activity incurs, and the amount of product that is fed or consumed. The financial analyst, however, is concerned with development of budgetted average product costs. He must roll the costs of feed and intermediate products down to the final stage products. This paper bridges these two interests by demonstrating a methodology for the direct translation of the solution of the linear programming model to the development of rolled product standard costs.

Vol. 4
1976

INTRODUCTION

THE APPLICATION of linear programming models to manufacturing processes has resulted in improved decision making procedures for operating and planning industrial facilities. The question of using arbitrary cost allocation schemes to determine most profitable levels of production has been avoided by examining each alternative production activity in terms of the added production cost the activity incurs and the amount of product it produces and/or consumes. Thus, a linear programming formulation effectively decouples the cost relationships of sequences of production events and at the same time, through material balance relationships, recouples them into an integrated whole. The planning engineer concerned with specifying levels of production activity to the individual centers need not be concerned with a cost accounting formulation of average product costs to make his production decision. He is only concerned with added variable processing costs for incremental levels of operation. The cost account-

ant, however, is concerned with both historical and budgeted average product costs. This paper will demonstrate a methodology to bridge these two interests.

BACKGROUND

In a typical manufacturing complex such as may be found in the chemical process or oil industry, there are a number of material balance activities. Each activity has a material balance effect on the amounts of products available in a plant-product inventory pool; thus, a purchase activity increases the amount of product available, a manufacturing activity will generally decrease some product inventories, while increasing other product inventories, and interplant transfers will increase inventory at one location at the expense of the other's. In addition to having an effect on the product pools, material balance activities will generate added cost or revenues.

Manufacturing sets of activities are typified by a number of interdependent unit processes where a product produced from one unit is fed into another unit, thus creating a chained network of flows. The planning or budgeting cycle of such a firm consists of product demand and prices estimated by the Marketing Department, production efficiencies, variable costs of production, usages and capacities determined by the Manufacturing Department, and purchasing and distribution costs estimated by the Supply and Distribution Department. The manufacturing planning engineer will formulate these material balances and variable cost relationships into a linear programming model of the company and will determine optimal levels of purchases, production, distribution, and sales. Following the development of this optimized detailed plan, the Cost Accounting Department will calculate the standard costs of each product produced. Because the interrelationships of these materials balances activities are often quite complex, the development of the standard costs for each product is time consuming. If midway in the planning cycle, the product mix or variable costs change due to changes in raw material prices, unit efficiencies, sales demands or sales prices, the entire process of standard cost calculation must be largely repeated. The ability to automate this process by having a direct linkage between the determination of optimal levels of production and standard cost calculation will shorten the planning cycle time dramatically.

ROLLED AVERAGE VARIABLE COSTS

As indicated earlier, in a production-oriented linear programming model an added variable cost is associated with the various material balance activities. Once an optimal-solution for the linear programming problem has been obtained, we wish to know the rolled average variable cost of producing each

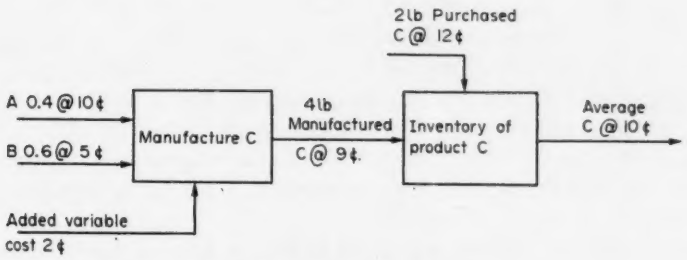


FIG. 1. *Rolled average variable costs.*

product, both intermediate as well as final product. In Fig. 1, for example, the rolled cost for manufacturing product C is $0{\cdot}4(10) + 0{\cdot}6(5) + 2$ or 9¢/lb. In addition, product C can be purchased for 12¢/lb. If 2lb of product C are purchased and 4lb are manufactured, then the rolled average variable costs for product C is $[9(4) + 12(2)]/6$ or 10¢/lb.

Substantially more complex interrelations can be obtained where there is a large network of product flows. The determination of the rolled costs is not available from the optimal solution in terms of dual variables or reduced costs since these numbers are only concerned with incremental changes from the optimal solution, whereas the rolled cost is an average cost considering the various production (purchase, transfer, etc.) alternatives at the production levels specified by the optimal solution. Since products can be produced at different plant locations as the result of multi-stage processes, the calculations of the rolled cost can be difficult if one tries to start with the raw material purchases and follow the job process stream. We have the chicken or egg problem here and the approach we have taken to calculate the rolled average variable costs is through the use of simultaneous linear equations derived from the linear programming model and its optimal solution.

Vol. 4
1976

LINEAR PROGRAMMING MODEL

The linear programming model to be considered here consists of material balance constraints (rows) imposed upon the different possible activities (columns); the activity levels being the unknowns. Thus, the rows correspond to the different possible product/plant combinations. The columns correspond to the different activities such as purchases, production, sales, substitutions, transfers, and exchanges (trades, conversions).[1] These activities fall into two (not mutually exclusive) groups: those activities which create a product (indi-

[1] For clarity the inventory activities will be excluded, but can be included if desired without loss of generality.

cated by negative coefficients) and those activities which consume a product (indicated by a positive coefficient). For example, purchases create a product at a specified plant while sales consume a product at a specified plant; production consumes certain products while creating other products at a specified plant; transfers transfer a product out of a plant (consumption) and into a different plant (creation), etc. Then, the material balance constraint for each product/plant row is that consumption cannot exceed creation or consumption–creation ≤ 0.

There is also a profit row which will not be considered part of the matrix but will be considered separately.

Notation

Let:

m	be the number of matrix rows corresponding to a product/plant.
n	be the number of matrix columns corresponding to a non-sales activity.
$P_i, i = 1, ..., m$	refer to the i^{th} product/plant.
$A_j, j = 1, ..., n$	refer to the j^{th} (non-sales) activity.
c_i	be the rolled average cost/unit of P_i, $i = 1, ..., m$ (to be determined).
T_i	be the total rolled average cost of P_i, $i = 1, ..., m$.
u_j	be the activity level of $A_j, j = 1, ..., n$.
v_j	be the added variable cost (from the profit row) for one unit of $A_j, j = 1, ..., n$.
V_j	be the total variable cost (added variable cost and cost of the direct feeds) for one unit of $A_j, j = 1, ..., n$.
$A = (a_{ij})$	be the $m \times n$ matrix of coefficients.
$P = (p_{ij})$ $= a_{ij}$ if $a_{ij} > 0$	be the $m \times n$ matrix of positive coefficients.
$= 0$ Otherwise	
$N = (n_{ij})$ $= -a_{ij}$ if $a_{ij} < 0$	be the $m \times n$ matrix of negative coefficients.
$= 0$ Otherwise	

Simultaneous linear equations

The total cost T_i of product/plant P_i is the sum of the costs of the materials used in creating it and the costs of the activities which create it. P_i is created by those activities A_j, $j = 1, ..., n$, for which a_{ij} is negative. The total cost of A_j is $u_j \cdot V_j$.

In a production unit, several products may be produced simultaneously (coproducts). For simplicity, we will distribute the costs of production evenly among the coproducts produced, so that every unit of product produced has the same cost allocated to it. We do this by dividing the total cost of production

by the total number of units produced. Thus, the total cost of A_j allocated to P_i is $(u_j \cdot V_j)\, n_{ij}/s_j$ where[2]

$$s_j = \sum_{k=1}^{m} n_{kj}.$$

The total cost T_i for P_i due to the activities which create it is

$$T_i = \sum_{j=1}^{n} n_{ij} u_j V_j / s_j \quad i = 1, \ldots, m. \tag{1}$$

Now V_j has two cost components. The added variable cost is v_j. The feed cost is $a_{kj} \cdot c_k$ for those product/plants P_k for which a_{kj} is positive. Hence,

$$V_j = \sum_{k=1}^{m} p_{kj} \cdot c_k \quad j = 1, \ldots, n.$$

Therefore,

$$T_i = \sum_{j=1}^{n} n_{ij} u_j (v_j + \sum_{k=1}^{m} p_{kj} \cdot c_k)/a_j \quad i = 1, \ldots, m. \tag{2}$$

Vol. 4
1976

The total units of P_i created by A_j is $n_{ij} \cdot u_j$ and hence

$$T_i = c_i \sum_{j=1}^{n} n_{ij} u_j \quad i = 1, \ldots, m.$$

By substituting and rearranging, we get

$$c_i \sum_{j=1}^{n} n_{ij} u_j - \sum_{\substack{k=1 \\ k \neq i}}^{m} c_k \sum_{j=1}^{n} p_{kj} n_{ij} u_j / s_j = \sum_{j=1}^{n} \frac{n_{ij} u_j v_j}{s_j} \tag{3}$$

for all $i = 1, \ldots, m$,

recognizing that $n_{ij}\, p_{ij} = 0$. We now have m equations in the m unknowns, $c_i, i = 1, \ldots, m$.

The set of m equations given by (3) can be written in matrix notation as follows:

Let $C = (c_i)$ be the $m \times 1$ vector of product/plant average rolled costs.
Let $D = (d_{ik})$ be a $m \times m$ matrix so that:

$$d_{ik} \equiv \sum_{j=1}^{n} n_{ij} u_j \qquad \text{if } i = k \qquad \text{for } i = 1, \ldots, m,$$

$$d_{ik} \equiv -\sum_{j=1}^{n} p_{kj} n_{ij} u_j / s_j \qquad \text{if } i \neq k \qquad \text{for } i = 1, \ldots, m \text{ and } k = 1, \ldots, m.$$

[2] $s_j \neq 0$ since after eliminating sales activities all remaining activities contain at least one negative coefficient.

Let $E = (e_i)$ be a $m \times 1$ vector such that

$$e_i \equiv \sum_{j=1}^{n} \frac{n_{ij} u_j V_j}{s_j} \qquad i = 1, ..., m.$$

Then, $D \cdot C = E$ or $C = D^{-1} E$.

D and E can be calculated as follows:

Let

$U = (u_j)$ be a $n \times 1$ vector of activity levels;
$U^* = (u^*_{ij})$ be a $n \times n$ matrix such that $u_{jj} \equiv u_j$ and $u_{ij} \equiv 0$ for $i \neq j$;
$W = (w_j)$ be a $n \times 1$ vector such that $w_j \equiv v_j/s_j$;
$U' = (u'_{ij})$ be a $n \times n$ vector such that $u'_{ij} \equiv 0$ $i \neq j$;
$u'_{jj} \equiv u_j/s_j$; and

$\overline{NU}$ be a matrix formed such that its diagonal elements are the elements of the vector NU and whose off-diagonal elements are zero. Thus $D \cdot C = E$ may be rewritten as $(\overline{NU} - NU'P^t)\, C = NU^*W$, (where P^t is the transpose of P). We now have a set of simultaneous equations whose solution yields C, the vector of unkown cost coefficients c_i.

Illustration

A company operates two production units (see Fig. 2). A and B are purchased for 8¢/lb and 5¢/lb respectively. D and E are sold for 12¢/lb.

Let

x_1 be the amount of A to be purchased,
x_2 be the amount of B to be purchased,
x_3 be the amount of C to be made in unit 1,
x_4 be the amount of E to be made in unit 2,
x_5 be the amount of D to be sold,
x_6 be the amount of E to be sold.

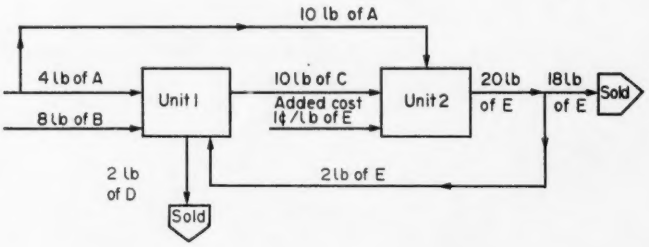


FIG. 2. *Rolled cost illustration.*

Then the above example can be written as a linear programming problem where the five inequalities correspond to the material balance constraints on the five products A, B, C, D and E.

$$
\begin{array}{lrrrrrrrrrl}
\text{Maximize} & Z = & -8x_1 & -5x_2 & & & -x_4 & +12x_5 & +12x_6 & & \\
\text{Subject to} & A & -x_1 & & & +0{\cdot}4x_3 & +0{\cdot}5x_4 & & & & \leq 0 \\
& B & & -x_2 & & +0{\cdot}8x_3 & & & & & \leq 0 \\
& C & & & & -x_3 & +0{\cdot}5x_4 & & & & \leq 0 \\
& D & & & & -0{\cdot}2x_3 & & +x_5 & & & \leq 0 \\
& E & & & & +0{\cdot}2x_3 & -x_4 & & +x_6 & & \leq 0
\end{array}
$$

According to our notation, $m = 5$ and $n = 4$.

$$
P = \begin{bmatrix} 0 & 0 & 0{\cdot}4 & 0{\cdot}5 \\ 0 & 0 & 0{\cdot}8 & 0 \\ 0 & 0 & 0 & 0{\cdot}5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0{\cdot}2 & 0 \end{bmatrix} \qquad N = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0{\cdot}2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad U = \begin{bmatrix} 14 \\ 8 \\ 10 \\ 20 \end{bmatrix}
$$

$$
U^* = \begin{bmatrix} 14 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} \qquad \begin{array}{ll} v_1 = 8 & s_1 = 1 \\ v_2 = 5 & s_2 = 1 \\ v_3 = 0 & s_3 = 1{\cdot}2 \\ v_4 = 1 & s_4 = 1 \end{array} \qquad W = \begin{bmatrix} 8 \\ 5 \\ 0 \\ 1 \end{bmatrix}
$$

$$
NU^* = \begin{bmatrix} 14 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} \qquad E = NU^*W = \begin{bmatrix} 112 \\ 40 \\ 0 \\ 0 \\ 20 \end{bmatrix}
$$

$$
U' = \begin{bmatrix} 14 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8{\cdot}33 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} \qquad NU' = \begin{bmatrix} 14 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8{\cdot}33 & 0 \\ 0 & 0 & 1{\cdot}67 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix}
$$

$$
NU'P^t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 3{\cdot}33 & 6{\cdot}67 & 0 & 0 & 1{\cdot}67 \\ 0{\cdot}67 & 1{\cdot}33 & 0 & 0 & 0{\cdot}33 \\ 10 & 0 & 10 & 0 & 0 \end{bmatrix} \qquad NU = \begin{bmatrix} 14 \\ 8 \\ 10 \\ 2 \\ 20 \end{bmatrix} \qquad NU = \begin{bmatrix} 14 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 20 \end{bmatrix}
$$

$$
D = \begin{bmatrix} 14 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ -3{\cdot}33 & -6{\cdot}67 & 10 & 0 & -1{\cdot}67 \\ -0{\cdot}67 & -1{\cdot}33 & 0 & 2 & -0{\cdot}33 \\ -10 & 0 & -10 & 0 & 20 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} 0{\cdot}0714 & 0 & 0 & 0 & 0 \\ 0 & 0{\cdot}125 & 0 & 0 & 0 \\ 0{\cdot}0325 & 0{\cdot}091 & 0{\cdot}109 & 0 & 0{\cdot}0091 \\ 0{\cdot}0325 & 0{\cdot}091 & 0{\cdot}0091 & 0{\cdot}5 & 0{\cdot}0091 \\ 0{\cdot}0519 & 0{\cdot}0455 & 0{\cdot}0545 & 0 & 0{\cdot}0545 \end{bmatrix}
$$

$$
C = D^{-1}E = \begin{bmatrix} 8 \\ 5 \\ 7{\cdot}462 \\ 7{\cdot}462 \\ 8{\cdot}723 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 82/11 \\ 82/11 \\ 96/11 \end{bmatrix}
$$

Vol. 4
1976

SOLUTION

No attempt to determine mathematically when this set of simultaneous linear equations has a solution or whether or not this solution is unique has been made. However, two facts should be mentioned.

In order to solve this set of simultaneous linear equations, the value of c_i, $i = 1, ..., m$, must be defined. However, if product/plant P_q was not created, then c_q is not defined. In this case, we must eliminate the equation corresponding to P_q. Furthermore, if P_q was not created, it cannot be consumed so that, in the other equations, the coefficient of c_q will be zero. Thus, we can eliminate the row and column corresponding to P_q.

The second consideration is whether or not the model includes recycling. As mentioned previously, recycling occurs when some product is used to 'create' one of its own precursors (possibly several stages removed).

Another example of recycling is when product A is used to make product B and product B can then be exchanged for product A at some ratio possibly with some exchange of money. When there is no recycling, one can be sure that there is a unique solution since one can calculate, starting with the raw materials, the rolled cost at each step in the process stream.

When the model includes recycling, there may not be any solution. For example, a pound of product A is used to make a pound of product B at a cost of 5¢/lb. Thus, $c_B = c_A + 5$. At the same time, a pound of B can be exchanged for a pound of A and 10¢ so that $c_A = c_B - 10$. Clearly, these two equations have no solution. In the same way, one can construct cases of multiple solutions (in which case arbitrary assignments can be made). From a practical standpoint, however, neither of these cases have been encountered. Once those products which were not created were removed from consideration, a unique solution has always been obtained.

COST ALLOCATION

When two or more products are produced in a production unit, it has been assumed that the total cost (feed cost and variable cost) was divided evenly among the products produced so that the rolled cost per pound of product produced was the same for all the products. In certain cases, one may wish to associate the total cost with only one product or one may wish to divide the total cost by some other formula. In some instances firms allocate costs of coproducts according to the net sales prices of the products and in other instances allocation in a petrochemical operation among coproducts was accomplished in proportion to the carbon content of the coproducts. Different cost allocation schemes can be handled by changing equations (1) and (2); that is, replace n_{ij}/s_j by t_{ij} where t_{ij} is the percentage of the total cost of A_j allocated to P_i.

Applications and extensions

In many companies, the Financial Department calculates the cost of operating a production unit, showing the individual component costs on a per pound as well as total poundage basis. Using the calculated rolled costs, the *LP* matrix and its solution, these component costs can be calculated quickly and easily. Similarly, the rolled cost can be used to calculate the economic profit [(sales price—rolled cost) × activity level] for each sale and exchange.

An interesting variation on the rolled cost model is to calculate the percentage of a product's cost attributable to a given raw material. This problem can be easily handled by taking the *LP* matrix and its solution and eliminating all added variable costs (for all activities including purchases) except for the selected raw material. Then, only the raw material's purchase cost will be rolled through the system so that the resultant rolled costs are attributable to this raw material.

An obvious extension of this paper is the inclusion of fixed costs in the rolled cost calculations. If fixed costs are associated with a given production operation, the equation can easily change to handle them. In fact, the total cost of activity A_j is now $u_jV_j + F_j$ where F_j is the fixed cost associated with activity A_j, so that equation (1) is now:

$$T_i = \sum_{j=1}^{n} n_{ij}(u_jV_j + F_j)/s_j \quad \text{for } i = 1, \ldots, m.$$

Carrying this down through equation (3) gives:

$$c_i = \sum_{j=1}^{n} n_{ij}u_j - \sum_{\substack{k=1 \\ k \neq i}}^{m} c_k \sum_{j=1}^{n} P_{kj}n_{ij}u_j/s_j = \sum_{j=1}^{n} (F_j + u_jv_j)n_{ij}/s_j$$

for $i = 1, \ldots, m$.

Programming considerations

With the capability of advanced Matrix Generators and Report Writers such as GAMMA 3 (Bonner and Moore's processor) matrices can be easily generated and output from the solution files to the LP can be easily obtained. Additional programs can be written linking this output to systems which solve simultaneous equations such as IBM's scientific subroutine package. If the matrices are large, the use of IBM's MATLAN, a Matrix Language System which only stores non zero coefficients of matrices to reduce storage requirements has been found to be most effective.

Computer programs to calculate rolled costs have been implemented on the IBM 370/165. For a problem with 931 rows (product/plant combinations of which 468 were active) and 1760 columns (1553 non-sales activities), 5·63 CPU minutes were required to put on tape the linear programming matrix including the profit row, the column activity levels, text information describing the pro-

Vol. 4
1976

duct/plants and sales information, using GAMMA 3 and MPSX. This tape was processed by a MATLAN program which took 3·70 CPU minutes to read in the above information, set up the 468 simultaneous linear equations, solve the equations by inverting a 468 × 468 matrix, and print the rolled costs and an economic profit report.

CONCLUSION

The problem of efficiently calculating rolled product standard costs for complex interrelated manufacturing enterprises was defined. The need and desirability of linking the calculations of these standard costs to the optimized production plan provided by an LP model was described. A methodology for the direct translation of the solution of the LP to the calculation of the rolled product standard costs using matrix algebra was developed, illustrated, and computational experience noted.

ADDRESS FOR CORRESPONDENCE: *Dr. Michel Araten, Vice President, Chase Manhattan Bank NA, 1 Chase Manhattan Plaza, New York, NY 10015, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Effect of Telecommunications on the Location of Office Employment

ROGER PYE

(*Received September 1975; in revised form January 1976*)

The face-to-face communications of office employees are classified and assessed for their suitability for transfer to new telecommunications media. This assessment is based on a series of psychological experiments and studies of users' attitudes. Large scale surveys of meetings enable the proportion which can be transferred to new media to be estimated.

If this transfer were to occur, it might encourage the decentralisation of offices or departments from central business districts. The extent of this encouragement is considered, with the possibility of even more fundamental changes in office employment, such as working at home or in work centres near the home.

Vol. 4
1976

INTRODUCTION

In England, Scotland and Wales, office employment is heavily concentrated in Central London. Outside the Greater London Council area the number of office jobs per 100 population over the age of 15 ranged smoothly, in 1966, from 6 in the most rural areas to 16 in the most industrial. But in the Greater London area the density was 25 jobs per 100 population. Moreover, most of these jobs in Greater London are concentrated in the conurbation's centre: the density in the remainder of London is only 12, the same as that in the surrounding Outer Metropolitan Area. The same imbalance can be seen from the percentages of population and office jobs in these areas which are given in Table 1. This heavy concentration of office employment results in shortages of

Table 1. Percentages of population and office jobs in England Scotland and Wales concentrated around London

Area	% of population	% of office jobs
S.E. England	20	43
Greater London	15	28
Conurbation Centre	0·5	14

suitable staff and accommodation in Central London. As a result, higher salaries must be paid to staff to induce them to travel, often considerable distances, to the central business district. Despite these higher salaries shortages remain. Similarly office rents are very much higher (often now £10 per annum per square foot higher) in Central London than in the rest of the country. Furthermore, it has been necessary to restrict, by office development permits, the increase in the amount of office space in the South East of England and more especially Central London. Hence, when an organisation is forced to move its offices, usually because its office space either is no longer sufficient or available, the combined pressures of shortage of space and high costs require it to consider relocation from London and indeed many firms have moved out. A recent major study has resulted in plans to decentralise over 30,000 senior civil servants from London to other metropolitan areas. During the period 1963 to 1974 the Location of Offices Bureau (LOB) recorded the decentralisation from London of 1615 organisations involving a total of 133,248 jobs, about 10% of the total number of office jobs in the London area in 1966. But as can be seen from Table 2, the great majority of these non-Civil Service moves are over comparatively short distances. How does this existing dispersal situation compare with common prophecies that telecommunications will result in the decline of central business districts, offices and even the city itself by enabling office workers dispersed throughout the land to communicate with each other electronically either from their own homes or from communications centres near their homes?

Vol. 197

TABLE 2. THE DISTANCES MOVED BY LOB CLIENTS, 1963 TO 1974

Distance from Central London	No. of moves	% of moves	No. of jobs	% of jobs
Inside GLC area	676	42	47,810	36
Beyond GLC area				
19 miles	135	8	6,665	5
20–39 miles	388	24	27,687	21
40–59 miles	104	7	10,930	8
60–79 miles	117	7	16,912	13
80 miles and over	195	12	23,264	17
Total	1,615	100	133,248	100

FIVE TYPES OF CHANGE

In the long term, offices could change in five ways:

(a) They could be relocated as complete units with no change in their internal structure. External contacts would be made via a public network of advanced telecommunications devices.

(b) Some departments, presumably those with little external contact, could be fragmented and decentralised. Contact between the fragmented employees and those remaining in the central business districts would be by private telecommunications systems.

(c) Offices might cease to be structured in departments. Instead, an organisation would own or lease a number of work centres throughout (one or more) metropolitan region(s), and its employees would be dispersed to the centre nearest their homes. The centres would be equipped with telecommunications and computing devices to enable contact to be maintained.

(d) Whereas in (c), employees were dispersed to centres occupied only by employees of their own organisation, they might alternatively be diffused to a greater number of neighbourhood work centres. An employee would attend the neighbourhood work centre nearest his home where he would work alongside employees of other organisations, with whom he might have no functional contact.

(e) Office staff might cease to work in any kind of office building but would work in their homes which would be equipped with communications facilities. The 'Relocation of Government' review [3] assumed that the principal disbenefit resulting from dispersal was the damage to Civil Service communications. Similarly three studies conducted for the LOB[1] show that the need to maintain communications is an important factor in office location decisions taken by private sector managers. This importance is supported by research conducted by Thorngren [7] in Sweden and Goddard and Morris in the UK [2] on the effects of relocation on patterns of office communication and on differences in patterns between offices which are intending to relocate and offices which have decided to remain in the capital.

Other factors are, of course, also important, especially the need to recruit and retain staff and the availability of office space.

Vol. 4
1976

HOW OFFICE WORKERS SPEND THEIR TIME

A number of studies have been made, by other researchers, of the ways in which managers spend their time. They have all produced quite similar results, and those produced by Stewart [6] are fairly typical (see Tables 3 and 4). Fifty per cent of their time was spent communicating with other people, even excluding dictating; only 33% of their time was spent alone. Furthermore most of the time spent with other people was spent with people in the same organisation. Thus from the point of view of communications, organisational changes which

[1] Can be obtained from the Locations of Offices Bureau, London.

TABLE 3. HOW MANAGERS SPEND THEIR TIME [5]

	%
Writing, dictating or reading company material	26
Figurework	8
Reading material originating outside the company	2
Informal discussions	43
Committees	7
On the telephone	6
Inspecting	6
Social activities	4

TABLE 4. TIME SPENT WITH OTHER PEOPLE

	%
People in the same organisation	
Immediate subordinates	26
Immediate superiors	8
Colleagues	12
Fellow specialists	8
Others	5
People in other organisations	
Customers	5
Others	6
Alone	30

keep work groups together will be easier to establish than those which split them up. Similarly if a public network of communications facilities were available then it would be cheaper (but not necessarily equally acceptable) to completely decentralise and maintain only external contacts than partially decentralise and maintain a greater number of internal contacts.

If dispersed, office workers would need to maintain two types of communication: those which are at present made interpersonally, i.e. face-to-face or by telephone; those made on paper, including that stored over periods of time (i.e. files, information systems, etc.).

INTERPERSONAL COMMUNICATION

Stewart found that the managers she studied spent about 50% of their time talking face-to-face and only 7% of their time on the telephone. The percentage of these contacts which could be conducted by telecommunications media, i.e. by video and/or audio plus facsimile systems, has been the subject of a considerable amount of research by the Communications Studies Group. The most recent method of estimating the percentage of contacts which could be transferred to these two types of system (video and audio plus facsimile) is based upon the Communications Studies Group's programme of laboratory experi-

ments and a survey of office communications. The laboratory programme has been supported primarily by Long Range Studies Divisions of Telecommunications Headquarters of the British Post Office and by the Civil Service Department, the department which conducted the study on which the recommendations for relocating Civil Servants is based. The programme has involved over 15 man years of research and over 2,500 subjects who have been asked to perform tasks face-to-face and by audio and video systems in order to measure the effects of medium of communication on the outcome of the task. Its findings can be summarised by the following extract from the final report to the contract sponsored jointly by Long Range Studies Department and the Civil Service Department which was completed in 1973.

> "Currently available forms of telecommunications, audio as well as television, would probably provide an effective substitute for many types of meeting which currently take place in Government and business, essentially those that are characterised by information exchange, co-operative problem-solving and routine decision making. On the other hand, for meetings characterised by bargaining or the development and maintenance of personal relationships audio systems provide an appreciably less effective and less acceptable medium than the face-to-face meeting. For the latter television systems are much more acceptable than audio, but although they are also more effective they still fall short of the face-to-face meeting. Finally, for meetings characterised by conflict the medium has a qualitative effect on outcome. This does not mean that audio is less effective than video and that the latter is worse than face-to-face, but this assumption may be necessary where caution is indicated."

Vol. 4
1976

The survey, also supported by Long Range Studies Divisions, and called the 1973 Office Communications Survey, asked respondents to complete up to 10 meeting record sheets. A total of 842 office workers recorded 3160 meetings. For each meeting they were asked the extent to which each of a number of functions and activities applied. (The question is shown in Table 5). This list of functions and activities had been derived from an earlier more detailed study,[2] On the basis of the rate to which these scales applied, and the results of the experimental work, meetings were allocated to the classes "suitable for a narrowband (i.e. audio plus graphics) medium", "suitable for a broadband (i.e. video) medium", "face-to-face necessary", and "experimental findings insufficient for allocation to be made". Details of this allocation are shown in Table 6. After weighting to remove certain statistical biases, it was estimated that the percentages of internal meetings and external meetings which could be substituted are as shown in Table 7. (Internal meetings were defined as ones which involved only staff from the respondents' own organisation and included at least one from outside his own department; external meetings involved people from outside the organisation).

[2] Unpublished report, can be obtained from the author.

TABLE 5. EXTRACT FROM MEETING RECORD SHEET

Q.9. A number of possible descriptions of the functions of the meeting and of what went on at the meeting are listed here. Please give each a score to show how well it applies to the meeting you are recording. If the description does not apply at all give a score of 0. If it applies completely give it a score of 6 (the maximum). It if applies to some extent but not completely, give it a score of 1, 2, 3, 4, or 5 as appropriate. Please give a score to each description.

Functions of the meeting:	
Delegation of work	..
Giving information to keep people in the picture	..
Information seeking	..
Discussion of ideas	..
Presentation of report	..
Disciplinary interview	..
Negotiation	..
Policy decision making	..
What went on at the meeting:	
Conflict	..
Problem solving	..
Forming impressions of others	..
Inspection of fixed object(s)	..
Any other description (Write in)	..

The most important aspects of these estimates are: the outcome of a substantial proportion of meetings would be affected if they were conducted by any telecommunications medium, and as such seem unsuitable for telecommunications substitution; the outcome of another very substantial proportion of meetings would not be affected and so seem very suitable for substitution; there are relatively few meetings which seem suitable for transfer to video systems but not audio plus graphics systems.

In addition to facilitating relatively few more meetings than audio systems, video systems are also very expensive. Estimates of the cost of installing two-way switched video of broadcast quality to every home in the U.S. range from \$1.2 trillion (i.e. $\$1.2 \times 10^{12}$) to \$720 billion ($\$720 \times 10^{9}$). The smaller of these is about the same as the current US investment in all transport systems, electric power and the telephone system together. Furthermore, Dickson and Bowers [1] have questioned the ease with which AT&T could raise the \$20 billion estimated to be required to convert 1% of US telephones to Picturephones®: in 1970 AT&T absorbed \$3.5 billion of external funding each year, or about 12% of the total available in the US economy.

Video services could of course be more cheaply provided in public studios or to a limited network of buildings. Unfortunately they lose much of their economic attractiveness if travel to a studio is necessary. The cost of travelling was compared with the cost of using a video system (estimated at roughly the current cost of using the British Post Office's Confravision system) for the meetings surveyed in the 1973 Office Communications Survey. When travelling time was

TABLE 6. DECISION TREE FOR TYPE ALLOCATION OF MEETINGS

All meetings		Allocation
More than 12 participants	—yes→	face to face
↓ no		
High scores on:		
↓ Inspection of fixed object (D)	→	face to face
↓ Disciplinary interview	→	face to face
↓ Conflict	→	face to face
↓ Giving information to keep people in the picture (not problem solving, discussion of ideas, or information seeking) (F)	→	?
↓ Policy decision making (B + C)	→	?
↓ Presentation of report (H)	→	?
↓ Delegation of work (K)	→	?
↓ Forming impressions of others (O)	→	broadband
↓ Negotiation (I)	→	narrowband
↓ Discussion of ideas, + giving information to keep people in the picture, + problem solving, + information seeking (G)	→	narrowband
↓ Discussion of ideas (J)	→	narrowband
↓ Information seeking and problem solving and giving information to keep people in the picture (N)	→	narrowband
↓ Information seeking and problem solving (M)	→	narrowband
↓ Giving information to keep people in the picture and problem solving (E)	→	narrowband
↓ Information seeking (A)	→	narrowband
↓ No high scores (L)	→	face to face

TABLE 7. THE WEIGHTED PERCENTAGES OF MEETINGS SURVEYED BY THE 1973 OFFICE COMMUNICATIONS SURVEY WHICH ARE SUITABLE FOR TRANSFER TO TELECOMMUNICATIONS

Base	External meetings 1416	Internal meetings 1908
Face-to-face	29	22
Uncertain	28	35
Broadband	7	3
Narrowband	37	40

Vol. 4
1976

costed at four times the salary cost and video systems are in-house, telecommunications would have been cost effective for about 82% of the meetings; when half the participants had to travel for half an hour to a studio the percentage was 38%; when all the participants so travel the percentage was 3%.

TABLE 8. THE COST OF POINT TO POINT NETWORKS PROPOSED BY THE MITRE CORPORATION FOR WASHINGTON DC

Network	Length (miles)	No. of channels (6MHz)	No. of users	Cost (10^3)
Federal	10	60	54	417
Commercial	50	28	300–3000	952
Municipal	73	28	271	807
Higher educational	15	28	23	183

The cost of four point-to-point networks which were proposed by the Mitre Corporation for Washington DC is shown in Table 8. A pair of channels could be used to carry a two-way video communication. Thus point-to-point service can provide fairly cheap video over short distances. The capacity of Mitre's proposed systems to carry video (which was not the principal purpose) can be estimated on the basis of Stewart's and our findings—each manager would (if all contacts suitable for video but unsuitable for audio were conducted by video) result in about 1 hr of video communication each week. Hence if the system could be fully utilised the federal network could serve about 2000 such managers.

Even if video systems were a cost effective means of communicating, at least 25% and possibly as much as 57% of all existing face-to-face contacts external to the department would have to continue face-to-face. Because of this residual need to travel, the three most extreme types of organisational change (working in two types of work centre and working at home) could not be set up, except perhaps for those performing extremely routine tasks which do not require face-to-face contact, in a highly dispersed manner. The staff would all have to live within the same metropolitan area, say, to enable the residual travel to take place.

The cost effectiveness of relocating from Central London to another location, assuming that a certain percentage of existing contacts in Central London must be maintained, can be investigated by comparing the costs of maintaining contact with the financial benefits (savings in rent, rates or local taxes and salaries) which were shown by Rhodes and Kan [5] to be obtainable as a result of relocation. It is easy to see that relocation is cost effective if

$$\frac{\text{total number of person contacts that must be maintained with Central London}}{\text{number of people in department}}$$

is smaller than

$$\frac{\text{savings in rent, rates and salary as a result of relocating one employee}}{(\text{fares cost} + A \times \text{journey time}) \text{ for return travel between the two locations}}$$

where: A is the value of time

The contacts included in the total are

either visits made by members of the subject department to other departments or outside organisations

or visits made by other departments to the subject department.

A similar relation can be used to compare potential receiving locations. The model was applied using parameters estimated as follows:

1. Savings in rent from the Location of Offices Bureau Surveys (June and Autumn 1974)
2. Savings in staff salaries from the Alfred Marks Bureau Survey of Secretarial and Clerical Salaries (January 1975)
3. Journey costs and times by first class rail travel.

It was found that the above test statistic for a number of surveyed organisations was consistent with the present pattern of dispersal, i.e. it advised a movement out of Central London but not beyond the South East. If narrowband, i.e. audio plus graphics systems were in widespread use an organisation maintaining somewhat greater levels of contacts should make such a move, but moves over greater distances receive little encouragement, because substantially greater savings cannot be obtained by moving further.

The model does not show that decentralisation from London to another metropolitan area could never achieve a net saving, merely that it is seldom worthwhile with the existing financial savings and communication costs, which might of course change in the future. Similarly different spatial distribution of office employment might enable more contacts to be replaced locally, thereby reducing communication costs. Typical spatial distributions of contacts maintained by relocated offices are shown in Fig. 1 (from Goddard and Morriss [2]).

LETTERS, MEMORANDA AND FILES

Less work has been done concerning letters and memoranda and access to files than interpersonal communication. The 1973 Office Communications Survey investigated the extent to which documents were involved in the meetings surveyed. Because questions were asked concerning the pre-arrangement of the meeting and the length of the documents it was possible to deduce whether

Vol. 4
1976

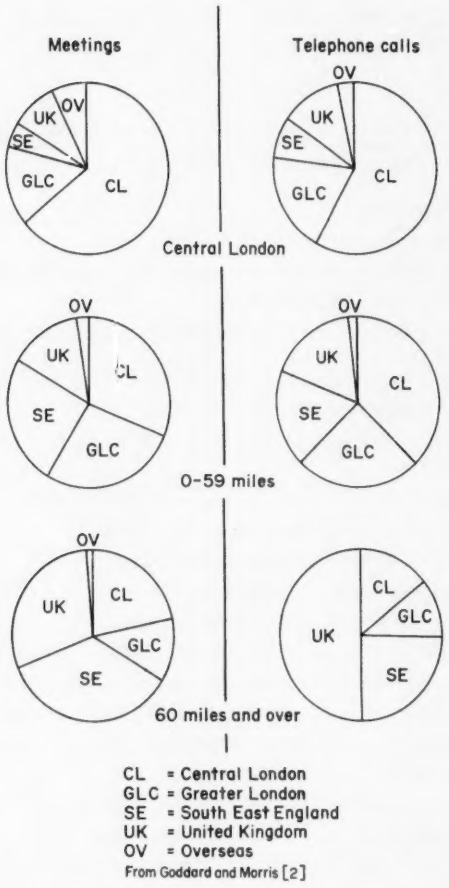


FIG. 1. *The proportions of meetings and telephone calls to different areas recorded by organisations in Central London and by organisations which had decentralised less than and more than 60 miles from London.*

or not the documents could have been distributed by post and 4-minute facsimile. The results are shown in Table 9.

A small survey (designed and analysed by the Communications Studies Group as part of the New Rural Society project funded by the Department of Housing and Urban Development) of a company in Connecticut found that the 25 managers and 4 secretaries involved received about 12 pages of correspondence per day from people who worked in another department or another building. Ninety-three per cent of this had been written recently; only 7% had

TABLE 9. THE WEIGHTED PERCENTAGES OF MEETINGS SURVEYED BY THE 1973 OFFICE COMMUNICATIONS SURVEY WHICH INVOLVE DOCUMENTS WHICH COULD BE SENT BY POST AND FACSIMILE WITHOUT ALTERING THE TIMING AND PREARRANGEMENT OF THE MEETING

Base	External meetings 1416	Internal meetings 1908
No documents used	48	61
Post feasible	13	9
Facsimile feasible	26	19
Facsimile not feasible	13	9

been obtained from files. Eighty-nine per cent of the documents were entirely alphanumeric in content and could be transmitted digitally. Correspondence of this volume could easily be handled by facsimile or on a data line, but not by telex, which would be too slow.

Alternatively CATV systems with limited feedback could be used to distribute such material to people working in their homes. A suitable system, the Electronic Information Handling system, was proposed by Mitre for Washington. They estimated a change of $18.50 per month per subscriber in addition to the charge for TV programmes. Also the federal, commercial and municipal networks they proposed were really intended to provide information systems to the offices. Each of the 6MHz channels, charged at $4 per hour, could be simultaneously used by about 300 people searching for and reading information. (Remember that two channels were needed for a single two-way video conference!)

Unfortunately the few trials which have taken place of remote information systems have not been well documented and little information exists on the extent to which staff access files held within their own departments.

All that can be said at this stage is that letters, memoranda and files do not seem to be an obstacle to relocation of complete organisations or complete departments. Too little is known to permit assessments of the feasibility of remote working in either the home or work centres.

SUMMARY

Since telecommunications are not suitable for all contacts, the extreme situation of work being dispersed on a geographically widespread basis to homes, work centres or even conventional offices, seems most unlikely. The residual travel is going to make much more likely dispersal through a metropolitan area, its suburbs and the immediately surrounding area. Given the existing cost differentials it is difficult to see that such dispersal would involve organisations moving from the London area to other major metropolitan areas. For this to be

Vol. 4
1976

possible they would have to be able to transfer many of their contacts to those areas and the great predominance of London as an office centre makes this increasingly unlikely. Perhaps the civil servants decentralised from London will act as growth points which will attract other offices. Alternatively, perhaps telecommunications can be used as a tool to make easier the transfer of contacts by building up contact communities away from the London metropolitan region which are as attractive as those which already exist in London. This would require telecommunications to be installed in these areas rather than as a radial network based on London.

Finally it may be thought that all of these arguments apply much more to countries such as England, France and Sweden, which only have one major office centre and not to those such as the USA which has several. The only difference is that such development will be regionalised within the USA. Given the existing trends of development, dispersal will continue and be encouraged to areas surrounding major metropolitan centres which will remain essentially independent and self contained.

REFERENCES

1. Dickson EM & Bowers R (1974) *The Video Telephone; Impact of a New Era in Telecommunications.* Praeger.
2. Goddard J & Morris D (In Press) *The Communications Factor in Office Decentralisation.* Pergamon Press, Oxford. (Progress in Planning Series, Vol. 6, Part 1.)
3. Hardman H (1973) *The Dispersal of Government Work from London.* HMSO Cmnd. 5322.
4. The Mitre Corporation (1972) Urban Cable Systems M72-57, The Mitre Corp.
5. Rhodes J & Kan A (1971) *Office Dispersal and Regional Policy.* Cambridge University Press, London.
6. Stewart R (1967) *Managers and their Jobs.* Pan Books (1970)
7. Thorngren B (1973) 'Swedish office dispersal' In *Office Location and Regional Development.* (Ed. Bannon M). An Foras Forbartha, Dublin.

ADDRESS FOR CORRESPONDENCE: *Roger Pye, Communications Studies Group, Joint Unit for Planning Research, University College, London, Wates House, 22 Gordon Street, London WC1.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Optimal Regional Distribution of Petroleum Products

CRAIG L MOORE

ANDRIS A ZOLTNERS

(*Received August 1975; in revised form February 1976*)

An economic model is presented which applies linear programming to determine the most efficient distribution of petroleum products among industries and regions. The model maximizes employment while (1) ensuring a balanced economy via input–output relationships, (2) maintaining at least the minimum production of basic goods and services necessary to prevent abnormally high inflation which could be caused by shortages, (3) ensuring that regional economies do not become severely depressed due to energy unavailability for their industrial base, (4) making sure that the industry capacity in each region is not exceeded, and (5) allowing policy makers control over the balance of payments. The model utilizes an input–output framework at both the national and regional level and makes adjustments for regional differences by incorporating regional multipliers.

Vol. 4
1976

I. INTRODUCTION

THE UNITED STATES has always enjoyed a relative abundance of national resources. Except for brief periods, usually associated with a war, there has never been a prolonged shortage of a basic commodity in America. During the few periods where shortages have occurred, the government has usually turned to price controls and rationing rather than relying on the market mechanism to allocate these resources.

Since 1972 there has been a growing concern over the potential for short-term shortages of petroleum. Production and refining capacity has not increased at a sufficient rate to meet swelling demands, and the 1973 Middle East war and Arab oil embargo against the industrialized West compounded this problem. In the long run, new sources of energy or improved technology may correct this situation, but many countries will face a continuing shortage of petroleum products, high prices and constraints on development for several years at least.

In 1973, the American government instituted price controls and established a

mandatory allocation program. This allocation plan was primarily voluntary and dealt only with broad aggregate categories of petroleum users. The formulation of a detailed distribution program, which would include both industrial and regional allocations, requires much more information than is presently available and the development of sophisticated tools foı economic planning and analysis.

This paper examines the potential use of linear programming in developing an oil allocation plan. The primary objective is to provide an economic model and a method of analysis which will help planners and policy makers determine the most efficient distribution of petroleum products among industries and regions in a nation.

A second objective is to develop an economic model for simulating the impact of a particular allocation program on the economy at both the national and regional levels. This is highly desirable because of the critical impact such programs have on prices, employment, production, and the basic quality of living.

While this general approach can be applied to almost any scarce resource, this paper will be limited to the petroleum allocation problem. This is done for two reasons. First, the distribution of petroleum products is a problem that will likely arise again. Petroleum serves as the backbone of energy production in most industrialized nations and, therefore, it is of critical importance to have the capability of formulating an allocation plan which will both optimize growth and maintain economic stability. Second, it is easier for most people to relate a model to a concrete problem rather than to deal with it entirely in abstract terms. The allocation of petroleum, therefore, serves as an illustration of a more general technique which could be applied to agricultural products, land, metals, water, or any other resource which is vital to society.

Vol.
197

II. A GENERAL DESCRIPTION OF THE MODEL AND ITS ASSUMPTIONS

The distribution process described in this paper distinguishes between three types of uses. The first category includes special allocations for the production of goods and services with a high social priority such as heating oil for hospitals, gasoline for police cars, or petrochemical feedstocks for manufacturing drugs. Home heating oil could be included in this category because it is a necessity for millions of American families. Allocations for these priority uses are made off the top of any existing supply of petroleum products and therefore are not included in the model.

Second, there are many goods and services (both private and public) which society must maintain on a minimum level. These would include final consumption goods such as clothing, agricultural products, consumer durables,

and certain governmental services, as well as the intermediate goods necessary for their production. In addition, a minimum level of production of many basic goods is essential to reduce inflationary pressures arising from shortages. Petroleum allocations for the production of all of these goods and services are insured through the incorporation of a set of constraints in the model.

The third category of petroleum products includes discretionary uses. All those goods and services not included in the first two groups would fall into this category. The model assumes that there are n groups or industries vying for this discretionary amount. These groups are based on the Standard Industrial Classification Code of the US Department of Commerce augmented by some additional special categories. They will be indexed by $i = 1,2,...,n$.

Since the model allocates spatially, users are further divided by region. Each industry has a different regional concentration of productive capacity. This not only causes variation in the regional demand for petroleum products, but can also create differences in the impact on employment between regions during a period of reduced output. Regions will be indexed by $j = 1,2,...,m$.

Vol. 4
1976

The model assumes either a predetermined level of available crude oil or predetermined levels of the four major refined petroleum products: fuel oil (also called distillates, heating oil), gasoline, heavy fuel oil (also called bunker fuel, residual fuel), and petrochemical feedstocks, identified by $k = 1,2,3,4$. If predetermined level of crude is assumed, then models are available for determining levels for the various petroleum products by establishing their market prices so as to equate refinery supply with consumer demand [2].

The LP model is a single period model. Typically, the demand for petroleum products and output levels for goods and services will be specified for a planning horizon which is flexible (it may be a week month, quarter, year, etc.).[1]

The decision variable in the model is output. Output is related to employment by a set of linear employment functions. The sum of these functions serves as the LP objective function. There are five constraint categories in the model. First, industry input–output relationships ensure a balanced economy. Minimum production levels are maintained so that the economy does not incur shortages of vital goods and to discourage supply based inflation. Another set of constraints assures that the available supply of refined petroleum products is not exceeded. Industry capacity constraints are included. Finally, regional economic constraints make sure that regional economies remain healthy.

The model

Maximize
$$\sum_{i=1}^{n} \sum_{j=1}^{m} M_j e_{ij} O_{ij}. \tag{1}$$

[1] Multi-period extensions of the model, incorporating inventory policies and lag effects remains the subject for future research.

Subject to:

(A) *Industry constraints*

$$\sum_{j=1}^{m} O_{ij} - I_i - \text{EXP}_i \geq \text{Demand}_i \quad (2)$$

$$\sum_{j=1}^{m} O_{ij} \geq \bar{O}_i \quad (3)$$

$$\sum_{j=1}^{m} O_{ij} - \sum_{h=1}^{n} \beta_{ih} \left(\sum_{j=1}^{m} O_{hj} - I_h - \text{EXP}_h - \text{Demand}_h \right) = 0 \quad \text{for } i = 1,2,\ldots,n \quad (4)$$

$$\bar{I}_i \leq I_i \leq \bar{I}_i \quad (5)$$

$$L_i \leq \text{EXP}_i \leq U_i \quad (6)$$

(B) *Supply constraints*

$$\sum_{i=1}^{n} \sum_{j=1}^{m} r_{ijk} O_{ij} \leq S_k \quad \text{for } k = 1,2,3,4 \quad (7)$$

(C) *Regional constraints*

$$M_{ij} \leq O_{ij} \leq C_{ij} \quad \text{for } i = 1,2,\ldots,n \quad j = 1,2,\ldots,m \quad (8)$$

$$M_j \sum_{i=1}^{n} e_{ij} O_{ij} \geq L_j (1 - \mu_j) \quad \text{for } j = 1,2,\ldots,m \quad (9)$$

where:

O_{ij} is the total output level for industry i in region j the planning period. O_{ij} is the decision variable.

e_{ij} is the employment/output ratio for industry i in region j. (This ratio and several others used in the model will be discussed in detail later.)

M_j is the employment multiplier for region j.

β_{ih} is the fraction of output in industry h that is consumed by industry i as a production input.

Note that $0 \leq \beta_{ij} \leq 1$ and $\sum_{i=1}^{n} \beta_{ij} = 1$.

I_i is the net change in inventory in industry i in the planning period.

EXP_i is the net import–export position for industry i in the planning period.

Demand_i is a target figure for aggregate demand for goods produced in industry i including government purchases, households, intermediate goods, capital formation, etc. (demand sector).

$\bar{O}_i$	is the minimal level of output required for industry i for the planning period.
$\bar{I}_i$, $\underline{I}_i$	are lower and upper bounds for net inventory change for industry i respectively.
L_i, U_i	are lower and upper bounds for the net import–export position for industry i respectively.
r_{ijk}	is the consumption/output ratio for industry i in region j for product k.
S_k	is the total supply of product k available in the planning period.
C_{ij}	is the output capacity for industry i in region j.
M_{ij}	is the minimum level of output for industry i in region j for the planning period. This may be positive if region j has a high dependence on industry i.
L_j	is the labor force in region j.
μ_j	is the maximum desired unemployment rate for region j.

The model maximizes employment. The objective function (1) is the sum of employment functions, $e_{ij}O_{ij}$, adjusted by a regional employment multiplier M_j, designed to reflect secondary and tertiary employment effects.[2] The employment functions relate the level of output, O_{ij}, to employment in industry i in region j. Linear estimations of these functions can be made with reasonable accuracy. Actual estimation of the employment functions, as well as estimation of other relationships used in the model, is deferred to the next section.

There are several reasons why the objective of the model is to maximize employment. First, the flow of household income is the mainspring of our national economy. If employment is not maintained, the economy faces a serious recession as well as a critical shortage of petroleum. There would be high social costs associated with widespread unemployment such as welfare payments and increased crime. At the same time, revenues from income taxes would fall and the government would face serious fiscal problems.

Constraints (2) ensure that each industry is able to meet final demand targets, by a combination of production, inventory manipulation, and imports. Government must have control over demand to prevent shortages of key commodities, retard recessionary pressures, and control inflation.

Demand$_i$ is a model input. It can be the estimated final demand for goods produced in industry i. Alternatively, these targets can be manipulated to ensure a level of supply which will retard inflationary pressure or will change the relative price mix of certain groups of goods and services (i.e. food, clothing, housing, etc.). Parametric linear programming could be used by policy makers

[2] M_j, the regional multiplier, equals $1/(1-kb)$ where kb is the propensity of regional residents to consume locally produced goods and services. kb can be estimated by the proportion of employment in the service sector of the region using a minimum requirements technique [1,3].

Vol. 4
1976

to evaluate the impact of various industry demand targets on total employment.

Constraints (3) stipulate that for each industry, minimal levels of output must be maintained in the production period. These minima are set at the national level to prevent discontinuities in production from period to period. Without this constraint the model could meet an inordinately high portion of demand from inventory and inputs in any given period. This may be undesirable.

Constraints (4) are the input–output equations necessary to strike a balance between final and intermediate goods at the national level. Without these constraints, the maximization of employment would be the only criterion for allocation and petroleum supplies would be channeled into an industry regardless of the demand for the output.

Constraints (5) restrict the allowable net inventory change. The national net inventory position for industry i is allowed to vary between upper and lower bounds. This is an important consideration if policy makers determine some margins of safety which should be maintained for key goods. Constraints (6) are trade restrictions. Net imports for industry i are restricted by upper and lower bounds. These constraints are added to insure that the nation maintains a reasonable balance of payments position. Such constraints would be set as a matter of national economic policy.

Four consumption functions, $r_{ijk}O_{ij}$, relate the average amount of petroleum product k required per unit of output in industry i in region j. These are assumed to be linear relationships. Constraints (7) reflect that available supplies, S_k are not exceeded.[3] The coefficient r_{ijk} is the consumption coefficient for petroleum product k in region j and industry i. Within an industry, this ratio primarily reflects climatic, technological, and scale variations between regions.

Constraints (8) are upper and lower bounds on the available production capacity in industry i in region j. These constraints are based on estimates of past output made by the Department of Commerce by region and SIC category.

[3] It should be noted that constraints (7) assume complementarity between fuel oil and heavy fuel oil. They do not assume substitutability. If one can be substituted for the other and vice-versa, then output O_{ij} has to be decomposed into O_{ij1} and O_{ij3} where O_{ij1} is that portion of output consequent on the use of fuel oil and O_{ij3} is that portion related to heavy fuel oil. Then constraints (7) should be replaced by

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ijk}O_{ijk} \leq S_k \qquad \text{for } k = 1,3$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ijk}O_{ijk} \leq S_k \qquad \text{for } k = 2,4$$

$$O_{ij} = O_{ij1} + O_{ij3} \qquad \text{for } i = 1,2,\ldots,n \quad j = 1,2,\ldots,m$$

If fuel oil and heavy fuel oil are substitutable and it is desirable to have one or the other, but not both, the model becomes an integer programming model.

The lower limit may be non-zero in key regional industries. This would be useful in protecting regional economic stability where the export base of the region was very dependent upon one or a few industries. Location quotients or an index of industrial concentration could be used in identifying such circumstances [1].

Constraints (9) set a lower bound on the employment of each region by applying a standard maximum unemployment rate, μ_i, to each region's labor force, L_i. The standard for maximum regional unemployment should be determined as a matter of policy. The regional employment multiplier, M_j, can be estimated by the size and economic composition of the region. The magnitude of the multiplier is inversely related to the proportion of the local economy which is involved in producing regional exports. Failure to include the multiplier for each region could result in a serious understatement of the lower employment bound. Initial unemployment, if allowed to fall to an unadjusted minimum, would only account for initial layoffs. The reduction in income would lead, in turn, to further rounds of unemployment due to a fall in regional income.[4]

The LP decision variable, output O_{ij}, can be viewed as the common denominator in the model. Both employment and petroleum product consumption are expressed in terms of output. In addition, regional economic constraints and input–output balance equations are set forth in terms of output. Once the optimal output levels have been resolved by the linear program, the associated allocation scheme and employment levels can be determined.

Vol. 4
1976

III. THE REGIONAL ASPECTS OF THE MODEL

One of the important features of the model is its focus on the regional economic impact of petroleum allocations. Both the objective function (1) and the supply constraints (7) include regional coefficients of production for each output category. The model does not include regional production of all inputs, but only for those factors of production which are directly involved in policy considerations. These inputs will be referred to as strategic factors of production and because of their importance in the model will be examined in detail.

a. *Regional-employment coefficients*

Regional employment coefficients, e_{ij}, are an integral part of the model, being used to relate industrial employment and output within each region. These labor coefficients may be estimated by computing total regional employ-

[4] The employment multiplier has the following effect. Suppose region j has a labor force $L_j = 100$, employment multiplier $M_j = 1{\cdot}5$ and a 10% initial unemployment due to an energy shortage. After secondary and tertiary effects are accounted for, the ensuing unemployment rate would be 15% (10% × 1·5).

ment within the planning period by a particular industry and dividing it by the output produced during that same period. Output can be measured in final dollars, physical units, or value added. Once this ratio is established, the employment coefficient will be determined.

b. *Regional coefficients for petroleum inputs*

The second set of strategic factors which are considered at the regional level are petroleum products. Regional-input coefficients for each petroleum product are included for every category of regional output used in the model. The inclusion of these coefficients will cause a shift in industrial allocations of these petroleum inputs from less efficient to more productive regions within the limits imposed by other constraints in the model. The upper bound on these shifts will be the production capacity for each industry within each region, and the lower bound will be a set of minimum output levels which must be maintained to insure national and regional economic viability. It is important to realize that the balance of the production mix is maintained, not at the regional level, but at the national level by a complete input–output module in constraints (4).

Regional coefficients for the four petroleum products can be estimated by the same method as described for labor coefficients in the previous section. The industrial output and petroleum consumption data required in the estimation process are not available at the regional level at the present time. This information would have to be collected by sampling the industrial sectors in each state. This method has been successfully employed in interregional trade research, but the coefficients generated did not break petroleum products into fine enough categories to be useful in this model [4, 5].

IV. SUPPLY LEVEL CONSIDERATIONS

Since the allocation LP is a single period model, the supply level for each refined petroleum product is the amount available for use in the planning period. This level depends on the quantity which is in inventory at the beginning of the period, the quantity produced during the period and available for consumption in the period, and the inventory required at the end of the period.

If estimates for the supply levels are available, they can be put directly into the model. Various alternative supply levels can be examined to determine their effects on the economy. Since an increase in the production of one refined petroleum product necessitates a decrease in another, the effects of various trade-offs can be analyzed. Model sensitivity analysis will be pursued in more detail in the section on alternative uses and strategies. If estimates for the supply levels are not available, they may be calculated by using an aggregate refining

LP model to equate supply and demand through the mechanism of market prices.[5]

V. ALLOCATION MATRIX

After solving the linear program, a solution matrix, $A_k = (A_{ijk}) = (r_{ijk}O^*_{ij})$, for each of the refined petroleum classes can be determined, where O^*_{ik} is the optimal solution. Once allocation A_{ijk} has been determined, allocation to individual firms or plants within industry i in region j can be made via some simple heuristics. For example, allocation can be based on the proportion of a firm's employment (or output) in the region's industry base.

VI. ALTERNATIVE USES AND STRATEGIES

The LP model is intended to be used as a planning tool. Various policy decisions can be tested to determine their effect on both regional and national employment and output. For example, effects of different minimum output levels for the various vital goods can be established. The disruptions resulting from emphasizing some goods over others can be determined. As mentioned previously, the effects of changing the refinery product mix can be ascertained. Foreign crude purchases can be analyzed on a cost–benefit basis. The outcome of increased capital investment and increased capacity can be estimated. In addition, the impact on the national economy of various regional constraints can be determined.

In addition to finding an optimal distribution, LP models have additional capabilities. Once an optimal allocation has been obtained, the effects of uncertainty can be examined via sensitivity analysis, particularly with respect to changes in the employment/output ratio e_{ij} and the consumption/output ratio r_{ijk}.

Parametric linear programming can be employed to determine petroleum requirements if the output of the economy were reduced by a variable percentage. Using this technique, the economy can be pared down incrementally to a point where a feasible supply of petroleum is available for an optimized output mix.

The model can be used in a number of other ways as well. It can be used to allocate a single refined petroleum product, e.g. fuel oil, many refined petroleum products, or it can be used to determine in what proportions crude should be refined into various petroleum products (which are in turn optimally allocated) to maximize employment subject to the numerous economic constraints mentioned earlier.

[5] We have received word that an aggregated refinery LP model has been developed by Bonner and Moore [6]. In addition, research in the determination of demand relationships for the refined petroleum products is underway [7].

Vol. 4
1976

Finally, the model can be used to evaluate the effects of alternative energy sources (e.g. coal, natural gas, solar heat, etc.). Alternative energy source levels are production inputs in the LP model. Changes in these levels correspond to changes in the input–output coefficients [constraints (4)].

VII. CONCLUSIONS

This model represents a potentially powerful tool which could be used to generate an optimal oil allocation plan. It has three distinct advantages over any method which has been developed up to this time.

First, it is flexible. It will allow administrators and planners to try various approaches to rationing by manipulation of the constraints on output and regional oil supply. Because it can be computerized, many different alternatives can be tried in a short period of time. The model can be run at different times throughout the winter with as many different planning horizons as the data base will allow. Second it has practical data requirements. It primarily uses standard regional economic data which is regularly collected by the Commerce and Labor Departments. The effort required to estimate regional petroleum-input coefficients does not seem unreasonable, and any alternative approach would have to employ fuel oil consumption data in some form. Both the machinery and expertise is available to obtain this information. Finally, the model makes explicit assumptions about priorities among users in real terms. Minimum levels of output and oil supply are assigned rather than an artifical indices or rankings.

Because of its realistic approach, practical data requirements, and flexibility, this model deserves serious consideration. It brings a powerful set of analytical and technological tools to bear on one of the most complex and vital problems that the United States has faced in recent times.

REFERENCES

1. Isard W (1973) *Methods of Regional Analysis*, Chapt. 7. M.I.T. Press.
2. Littlechild SC (1970) Marginal cost pricing with joint costs, *Economics Journal* **80**, 323–335.
3. Moore CL (1975) A new look at the minimum requirements approach to regional economic analysis. *Econ. Geogr.* **51** (4).
4. Polenske Karen R (1973) *The United States Multiregional Input–Output Model.* Heath Lexington.
5. Rodgers JM (1973) *State Estimates of Interregional Trade Flows.* Heath Lexington.

6. SISSON W (1973) Personal Communication, Marathon Oil Company.
7. The Gordian knot of gasoline prices. *Business Week*. December 15, 1973, p. 23.

ADDRESS FOR CORRESPONDENCE: *Dr. CL Moore, University of Massachusetts, School of Business and Administration, Department of General Business and Finance, Amherst 01002, Massachusetts, USA.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Two Issues in the Evaluation of Job-Shop Schedules

ISRAEL BOROVITS

PHILLIP EIN-DOR

(*Received October 1975; in revised form January 1976*)

Two issues are examined concerning the evaluation of job-shop schedules. First, machine and labor cost/utilization are suggested as alternatives to physical utilization criteria. Secondly, an exploration is made of the use of Kiviat Charts as a means of integrating multidimensional criteria for the overall evaluation of the quality of job shop schedules.

Vol. 4
1976

INTRODUCTION

THERE are numerous criteria for evaluating job-shop schedules. Le Grande [1], for example, lists ten such criteria. Table 1 contains eight of them, some in a slightly different form. This paper is an attempt to deal with two issues arising from the multiplicity of such criteria. The first concerns the measure of labor and machine utilization. The second arises from the multidimensionality of the criteria listed.

EVALUATING OVERALL QUALITY OF SCHEDULES

The considerable number of criteria available for measuring job-shop performance gives rise to the problem of integrating the multidimensional data into a single measure, so that alternative schedules or shop configurations may be compared. Le Grande, while comparing alternative scheduling rules, solved the problem by assigning a rank of 1 to that rule which appears best for a given criterion, with other rules being assigned ranks proportional to the degradation in their performance relative to the best rule. For each rule, the ranks it obtains on the set of criteria are weighted and summed so that the total relative rank is obtained for the rule.

There are two obvious problems with this kind of scheme. First, the ranking procedure for a given criterion may depend on the way it is defined. Thus one of Le Grande's criteria is percent of orders completed late: one might use instead its complement—percent of orders completed on time. In Le Grande's example, the rank order would not change as a result of the change in definition of the criterion. The rank values, however, would change considerably and their spread would be reduced from 52 to 27 percentage points. As a result, the overall rankings could well be affected.

Secondly, the weights assigned to criteria are largely subjective. Again, the final outcome may vary as a function of the weights chosen. A combination of these two problems may lead to the indication of a certain rule as optimal because it does very well on heavily weighted and well differentiated criteria, in spite of doing very poorly on low weighted, little differentiated criteria.

Between criteria expressible in monetary terms, the problems of ranking and weighting are solved by virtue of the single homogeneous criterion. It would thus appear to be preferable to use monetary transformations of criteria wherever possible. The first part of this paper suggests a way of doing this for labor and machine utilization. The second part deals with an additional tool for overall evaluation.

Labor and machine utilization

It appears to the authors that the cost of any input should weight the significance with which one reacts to information on its utilization. Thus the knowledge that a cutting machine whose cost is 2% of total machinery cost is utilized only 30% of the time is of different import than knowledge of the same degree of utilization with respect to a lathe whose cost is 25% of total machine cost.

This leads us to the concept of cost/utilization which applies utilization figures indirectly to the costs of physical inputs rather than to the inputs themselves. Cost is a common dimension which allows us to integrate utilization data for all the machines in a shop. We can then develop a single figure of cost/utilization for the entire shop. While it is impossible to develop meaningful figures of total machine performance directly from physical utilization, the common dimension of cost makes a single measure of merit both feasible and meaningful.

The cost/utilization factor measures the extent to which the outlay on all machinery is actually utilized. It is computed as

$$F_M = \sum_i p_i u_i$$

where p_i is the cost of the i^{th} machine in the shop as a percentage of total machine cost and u_i is the percentage utilization of the i^{th} machine. Thus F_M can vary from 0, in a job shop not utilized at all, to 1 in a perfectly utilized shop. Similarly, one can compute a cost/utilization factor for labor, F_L, where the

index denotes classes of labor. The method described below includes both computation of the cost utilization factor and its visual representation.

Cost/utilization: computation and representation

The graphic representation is of most interest for the machine inputs and will be demonstrated in this context. One may also graphically represent the labor inputs, if this is of interest. The computations in both cases are identical, as previously indicated.

The method is to construct a rectangular graph representing the maximum possible cost/utilization in the job shop. The horizontal axes of the graph represent percentage of total machinery cost contributed by each machine. The vertical axes represent percentage utilization of the machines. An histogram is constructed within this graph, the bars of which represent utilization of machinery cost and the area above the bars represents slack in the job shop. Once this histogram has been constructed, the cost/utilization factor is calculated, relating cost utilized to total machinery cost.

Example

In order to illustrate this method of cost/utilization analysis we have chosen a typical job shop that includes: two lathes, comprising 40% of total machine cost, four drills, accounting for 20% of cost, two cutting machines which absorb 30% and a press which accounts for the remaining 10% of cost.

Figure 1 exhibits four situations for this shop:

(a) The job shop is well balanced and well utilized. Balance is indicated by the fact that all bars in the histogram are about the same height. Near full utilization is indicated by the fact that all bars are close to their upper bound. The cost utilization factor in this case is 0·924.

(b) The job shop is well balanced but under utilized. The cost/utilization factor is 0·440.

(c) In this case the job shop is relatively well utilized but is somewhat unbalanced. The cost/utilization factor is 0·872.

(d) This job shop is both unbalanced and poorly utilized. One relatively cheap machine, the press, is an obvious bottleneck preventing utilization of the remaining machinery. Thus a small additional outlay of about 10% to double the capacity of this machine would permit a significant increase in system utilization. The cost/utilization factor is at present 0·500.

Once labor and machine cost utilizations have been computed, it is simple to compute overall cost utilization as

$$F = (C_M F_M + C_L F_L)/(C_M + C_L)$$

Vol. 4
1976

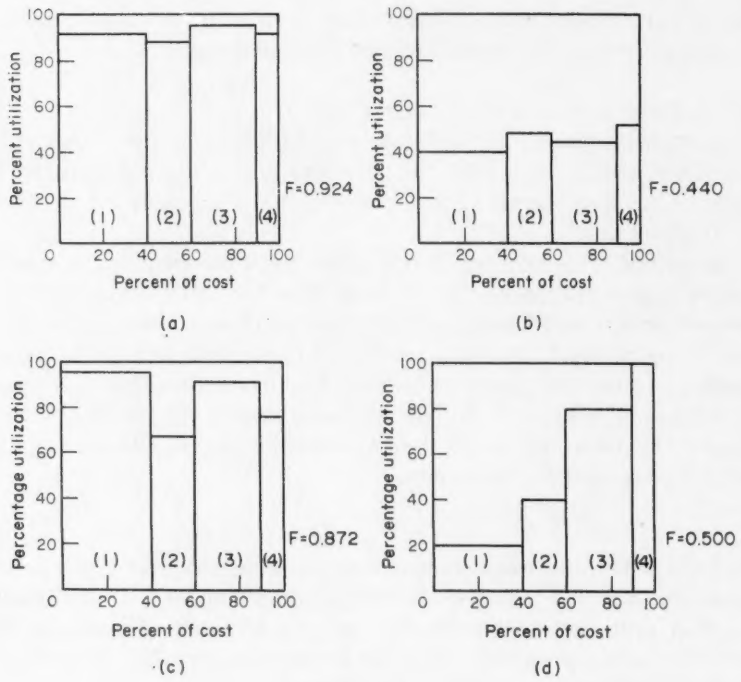


FIG. 1.

where F is the overall cost utilization factor for the shop (F_M and F_L are as previously defined, C_M is total machinery cost, C_L is total labor cost).

EVALUATION OF MULTIDIMENSIONAL CRITERION FUNCTIONS

As was pointed out above, the evaluation of multidimensional criteria by a weighting function may lead to the choice of schedules which do very well on some criteria and very badly on others. This same problem arises when the performance of computer systems is evaluated. An interesting attempt to provide an evaluation tool in this context is the Kiviat Chart, developed by Philip Kiviat [2]. These charts relate the various aspects of system performance, allowing one to see at a glance whether, in general, the system under consideration is well utilized or not. We believe that this same technique may be of value in evaluating job-shop schedules.

Kiviat Charts are prepared by first listing system parameters and the actual and ideal percentage utilization of the system with respect to each parameter.

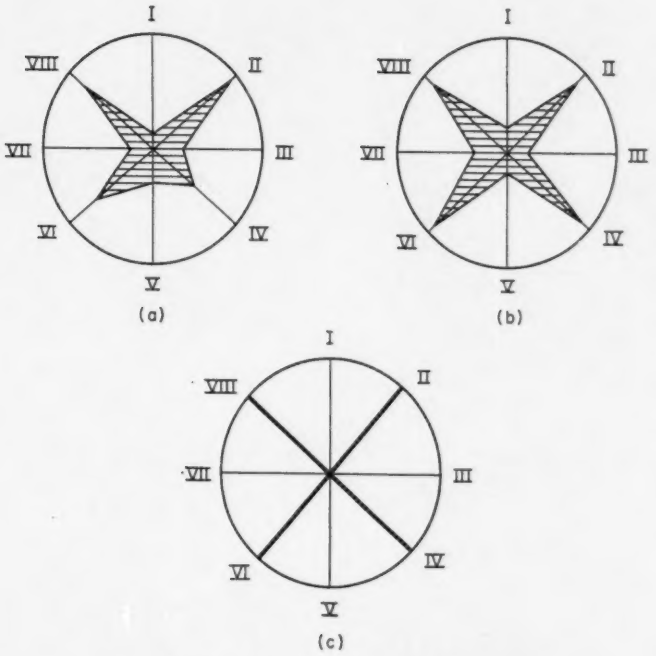


FIG. 2. *Examples of Kiviat Charts.*

Note that the number of parameters for which the ideal is 0% should equal the number for which it is 100%. Next, the parameter values are plotted on the radii of a circle, the center of which represents 0% and the circumference 100%. Each radius represents one parameter. Parameters with 0% and 100% ideal values alternate around the circle. The chart of Fig. 2a is typical of the result obtained when adjacent plots are joined.

In a well-utilized system, the shaded area is star shaped, with points close to the circle, e.g. Fig. 2b. In the ideal case, the shaded area is reduced to alternate radii of the circle, Fig. 2c.

Table 1 contains evaluation data for four of the scheduling rules examined by Le Grande. The data are adapted from Le Grande's paper and are exhibited in Kiviat Chart form as Fig. 3.

Vol. 4
1976

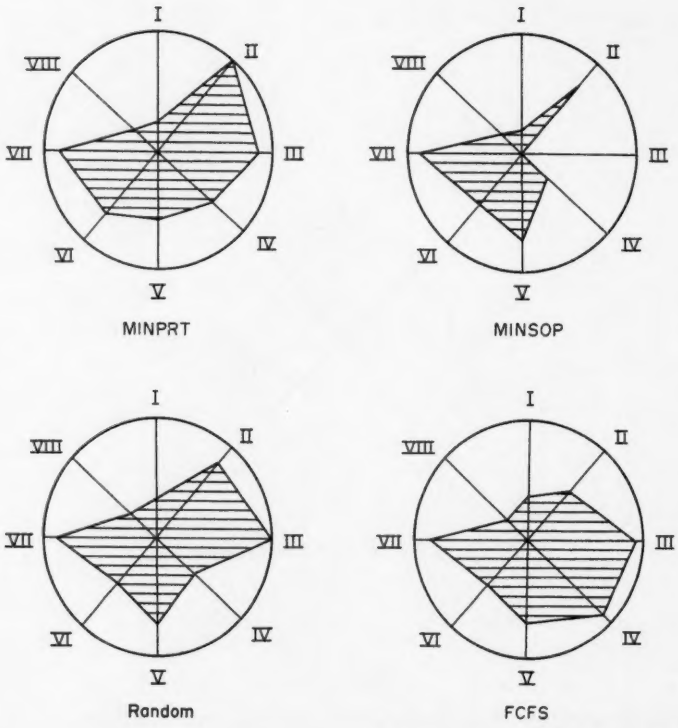


FIG. 3. *Kiviat charts evaluating performance of a job shop using four scheduling rules.*

TABLE 1. JOB SHOP EVALUATION DATA FOR FOUR SCHEDULING RULES

Criterion	Ideal	Performance Actual *			
		MINPRT	MINSOP	RANDOM	FCFS
I Per cent of orders completed late	0	25	20	30	38
II Rank of mean of distribution of completions	100	100	63	79	54
III Rank of standard deviation of distribution of completions	0	94	0	100	95
IV Average production time on orders/average total processing time	100	64	30	46	95
V Average per cent of orders waiting in shop	0	53	73	80	73
VI Labor (cost) utilization	100	63	58	58	59
VII Inventory carrying cost while waiting/total inventory cost in shop	0	93	92	93	93
VIII Machine (cost) utilization	100	27	25	25	25

* MINPRT = Minimum processing time per operation.
MINSOP = Minimum slack time per operation.
RANDOM = Random selection.
FCFS = First come first served.

The relative ranks of the scheduling rules were determined by Le Grande using both unweighted and weighted criteria. His findings were as follows:

Rule	Unweighted criteria		Weighted criteria	
	Relative rank	Rank order	Relative rank	Rank order
MINPRT	8·70	1	24·01	2
MINSOP	8·54	2	26·56	1
RANDOM	7·40	3	21·00	3
FCFS	6·93	4	19·20	4

It is clear from the charts that none of the scheduling rules performs very well when all criteria are considered simultaneously. It is also clear that different rules do better, or worse, in different areas.[1] Interestingly, all the rules do poorly on criteria VI through VIII, which deal with costs. Thus the scheduling rules examined, all of which are based on the time characteristics of jobs, do not perform very well in terms of economic criteria. It is quite likely that the use of scheduling rules with explicitly economic orientation might do better in this respect.

[1] All scheduling rules do badly on criterion VII, inventory carrying cost while waiting relative to total inventory carrying cost in the shop. It is possible that this criterion is unrealistic and requires reformulation.

Vol. 4
1976

SUMMARY

The concept of cost/utilization is an additional criterion aimed at improving the economic evaluation of job-shop scheduling rules. Its simplicity and ease of visual representation make it a readily applicable tool for operations management. The Kiviat Chart, applied in the context of job-shop scheduling, provides a simple means for integrating multidimensional criteria. It is thus of great help in analyzing and comparing the overall performance of scheduling rules.

REFERENCES

1. Le Grande E (1973) The development of factory simulation using actual operating data. *Mgmt Technol.* **III** (1).
2. Morris MF (1974) Performance evaluation: an emerging discipline of the computer science. Paper presented at the 2nd Jerusalem Conference on Information Technology.

ADDRESS FOR CORRESPONDENCE: *Dr. P Ein-Dor, Faculty of Management, The Leon Recanati Graduate School of Business Administration, Tel-Aviv University, PO Box 39010 Ramat Aviv, Tel-Aviv 61390, Israel.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

The Effective Use of Computer Resources

CW AXELROD

(*Received October 1975; in revised form February 1976*)

The distribution over time of jobs arriving for service at a computer installation will affect the turnaround time experienced by them, as does the sequence of service for jobs having arrived. However, if the pattern of delays due to congestion can be determined in advance, the submission times of jobs may be adjusted to increase the effectiveness of the system.

A theoretical model is developed in which users relate the net value of their jobs to submission time and turnaround time and the organization imposes submission and sequencing rules to increase the net gain of the system. A simulation is used to illustrate specific system environments and evaluation methods.

Vol. 4
1976

INTRODUCTION

THE 'COST per computation' for electronic digital computers has dropped dramatically over the past two decades [19, p. 323]. Software developments have enabled more efficient use of hardware, and sophisticated scheduling and pricing schemes have improved computer performance [9, 17, 20]. However, relatively little attention has been paid to the effective use of computers. Economists, generally interested in resource allocation, only focus on computer services occasionally [4, 21, 24]. In this paper we will explore methods of achieving increased computing effectiveness, assuming an efficient computer operation, noting that: "Efficiency is concerned with doing things right. Effectiveness is doing the right things" [6, p. 45].

We suppose an organization which has an internal computer installation.[1] The users are assumed to have some choice as to the number and timing of their jobs, in contrast to a fixed-schedule production environment.

THE COMPUTER USAGE ALLOCATION MODEL

Operations research literature contains many studies of computer job scheduling. Jobs are generally assumed to arrive due to exogenously-determined factors

[1] The case in which users can select external, as well as internal, computer services is considered in [1].

and the objective is to optimize some characteristic of the system, e.g. the minimization of turnaround time [5], lateness [7], or cost of waiting [10]. This decision process, through which the run sequence of jobs having arrived is determined, is termed 'micro-sequencing' to distinguish it from the process which governs the arrival sequence of jobs which we call 'macro-sequencing'.

Micro-sequencing is illustrated in the right half of Fig. 1, where the processing time of a particular job depends on the computer capacity, operating mode (e.g. multiprogramming, single programming, batch, time-sharing, distributed processing) and the running order (if sequence affects set-up time). Queueing (or waiting) time depends on the processing times of the jobs ahead, on the micro-sequence and the macro-sequence. The sum of processing time and queueing time is the turnaround time.

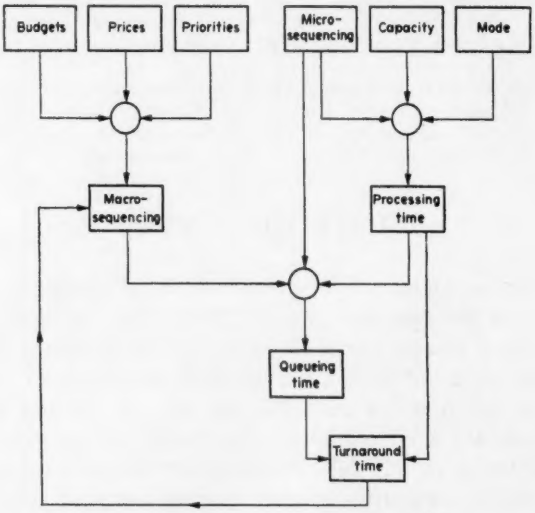


FIG. 1. *Computer usage allocation model.*

In this model, the level and timing of demand, i.e. the macro-sequence, is controllable by the organization.[2] From the left half of Fig. 1 it is seen that budget allocations, pricing rules and priorities are used to limit demand, both for individuals and in total. Users adjust the times at which they submit jobs (if at all) in response to expected turnaround time.

Micro-sequencing rules affect turnaround time, as shown above, and the time distribution of expected turnaround time affects the level and timing of demand. The stability of such a feedback system is considered elsewhere [1, pp. 57–70].

Prior work has not dealt explicitly with the timing aspect of demand and the interaction between micro- and macro-sequencing. In order to handle this

[2] In [16] the level of demand is controlled by tolls, but the timing factor is not considered.

important aspect of the model, we shall examine the relationship between the net value of a job and its submission time (i.e. time of day on a specific date) and expected turnaround time. Sharpe [19, p. 470] related value to completion time (i.e. submission time plus turnaround time), but did not consider the variation of value with each of the two components of completion time, nor did he examine the control of the timing of arrivals. It will be shown how the concept of a 'value hill' (i.e. value as a function of time of submission and turnaround time[3]) can be used to relate micro- and macro-sequencing and make the system more effective.

At this point we shall assume that job value can be both defined and measured. A brief review of the literature on these two topics is included in Appendix 1. Net value is gross value minus cost of processing, assuming no holding charge for a completed job [12]. Subsequent use of the word value should be taken to mean net value.

THE FORM OF THE VALUE HILL

The value hill for a job may be shown graphically as in Fig. 2. The contours join points of equal value and are called 'iso-value curves'. Each curve is labelled with its value (expressed in monetary units). For example if a job, submitted at time X, experiences a turnaround time of Y, its value at completion $(X + Y)$ is one unit.

Vol. 4
1976

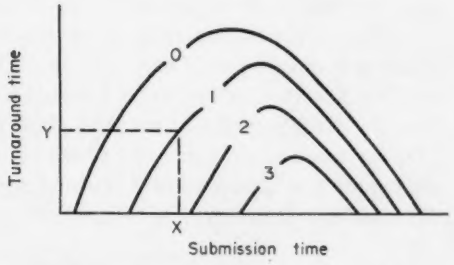


FIG. 2. *Value of job as a function of turnaround time and submission time (value hill).*

Since the value hill is single-valued the iso-value curves may not cross, although they can theoretically touch (indicating an instantaneous drop). Also, an iso-value curve cannot form a closed loop since value is non-increasing monotonic in turnaround time.[4] Otherwise value hills can take on a variety of shapes reflecting various sensitivities of jobs to time of submission and turnaround. In particular, the iso-value curves in Fig. 3 all have a minus-one gradient and job value is a function of completion time. For example, if the

[3] The relating of value to expected turnaround and submission is a simplification. Other factors, such as the variance of turnaround, also affect the quality of service.

[4] This assumes no holding cost. LeBon [12] examines the effect of a linear holding cost.

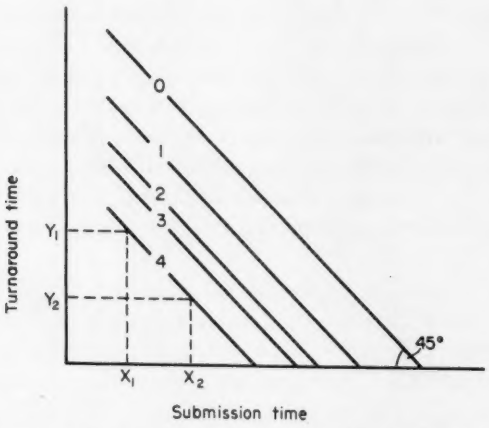


FIG. 3. *Value hill having the same value for the same completion time.*

job is submitted at X_1 or X_2 with turnaround times Y_1 and Y_2 respectively, the value is four units in either case and the completion times are the same, i.e. $(X_1 + Y_1) = (X_2 + Y_2)$. Sharpe [19, p. 470] implies this form of value hill by relating value to completion time. However, it is contended that a user may ascribe different values to the same completion time for a job submitted at different times; for example, an additional cost may be incurred for later submission due to, say, the need to use a messenger service versus internal mail or, perhaps, due to the effect of the psychological pressure of a closer deadline. Also, if a job is submitted at a specific time it may have no value if not processed by a certain time, but if it is known in advance that it would not be processed in time then general schedule changes might be made to allow for later running. An example would be the need to present a computer report at a periodic meeting, in which case the job would have to be completed, say, the prior day—otherwise it would be valueless. However, if it is known that the job will not

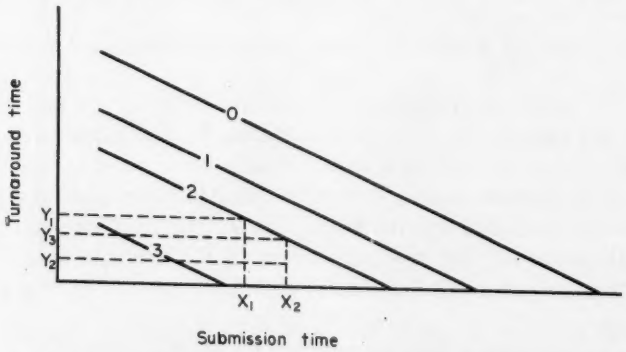


FIG. 4. *Value hill having different values for the same completion time.*

be available, then it may be possible to defer the presentation of the report to a subsequent meeting at which it may not have as much business impact or value. A simplified representation of such a value hill is shown in Fig. 4. Here $(X_1 + Y_1)$ and $(X_2 + Y_2)$ are again equal but the value realized is greater for (X_2, Y_2) indicating that the fall-off in value in relation to submission time is not as great as for Fig. 3. Conversely, submission at X_2 leads to the same value as (X_1, Y_1) if completion time is $(X_2 + Y_3)$, where Y_3 is greater than Y_2.

THE APPLICATION OF VALUE HILLS

The value hill can be used as the basis for both macro- and micro-sequencing.

Macro-sequencing

In macro-sequencing we assume that the user, if autonomous, will submit a job at a time yielding the highest value, but this does not necessarily produce the greatest total value for the system. The organization should try to obtain the maximum value for all jobs combined, as illustrated in Fig. 5. In Fig. 5a we see the value hill of Fig. 2 cut by the 'turnaround schedule', shown as AB, which represents the expected delay that a job would experience over time due to other jobs in the system.

Vol. 4
1976

The intersection of AB with the value hill produces job value as a function of submission time (Fig. 5b), with a maximum at X. Knowing value-hill functions for all jobs with positive net value during the submission period under consideration and the turnaround schedule (AB), one can develop the loss in value incurred by them if the job of Fig. 5a is submitted at all times within its feasible range. The loss-in-value curve shown in Fig. 5c is subtracted from the job-value curve of Fig. 5b to yield the net gain curve (Fig. 5d) which has a maximum at Y. Note that the submission times for maximum user value and maximum organizational value can be different (i.e. $X \neq Y$). One method of causing the user to submit jobs in accordance with the organization's optimization requirements is to charge users based on the value-loss function (which compares to peak-load pricing schemes [25]).

Micro-sequencing

Subsumed under the above analysis is a run sequencing discipline. Given that, at any point in time, jobs have arrived at the computer as a result of a macro-sequence determination, then the sequence in which the jobs should be run can be determined from the value hills. Note here that the micro-sequence affects the turnaround schedule, as shown in Fig. 1. If the value hills are simple in form, an algorithm may be developed, such as the c/t rule which will maximize the net gain if the value is linear decreasing in turnaround [14]. For the c/t rule jobs are processed in order of decreasing ratio of cost or value loss (c) to processing time (t). For complex value hills an heuristic approach must be used.

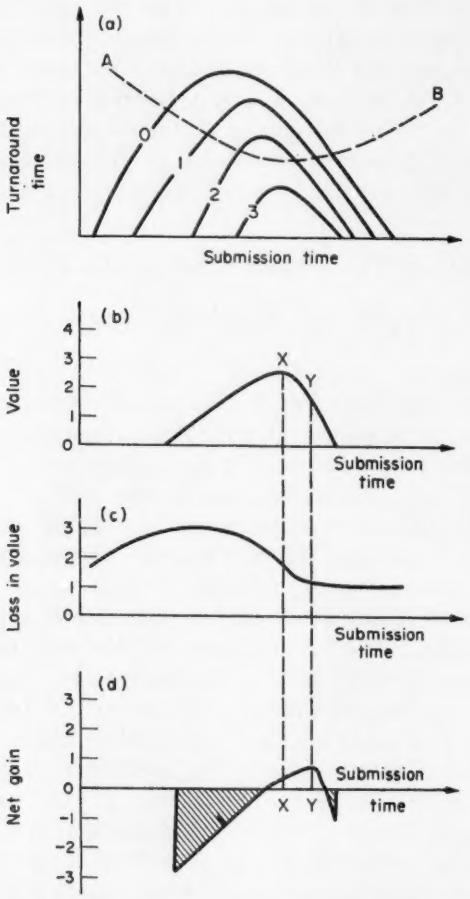


FIG. 5a. *Value hill cut by turnaround time schedule AB.*
FIG. 5b. *Job value as a function of submission time.*
FIG. 5c. *Value loss incurred by all other jobs in system as a function of the submission time of the above job.*
FIG. 5d. *Net gain obtained from the above job as a function of its submission time.*

SIMULATED SYSTEM

In Appendix 2 a simulated job stream is generated and passed through a series of sequencing programs using FCFS (first-come-first-served), SJF (shortest job first), priority and c/t micro-sequencing rules.

In one environment, jobs arrive at random and await service if the computer is busy. The value hills are as in Fig. 3, with jobs being submitted at the earliest available time. In the other environment, jobs arrive by appointment and are

served immediately upon arrival. These jobs have the value-hill form of Fig. 4. They do not have to wait for service, but may lose value due to later submission.

The two environments represent extremes, and are meant to be illustrative. The theoretical approach of the previous section falls between these two cases, namely, some measure of control of arrivals over which is superimposed a degree of randomness.

The system in both environments has the same total value for each of the density cases. It can be seen that the added flexibility of the extended job submission periods of the appointment environment allows for greater productivity and higher total value. It is also seen that the value-based micro-sequencing rules (i.e. priority and c/t) tend to produce higher value from fewer jobs run than the SJF rule. Note that the relative effectiveness of the rules as illustrated in Appendix 2 does not hold under all conditions and can vary with job parameters and densities.

CONCLUSION

This paper has laid a theoretical foundation which facilitates the evaluation of various macro- and micro-sequencing systems. In particular, the latter systems are related to each other explicitly, thereby creating a powerful job control mechanism. While the author appreciates the practical difficulties in obtaining job value hills, the literature reviewed in Appendix 1 provides many concrete suggestions regarding both the meaning and measurement of value. Furthermore, once the jobs' characteristics have been determined, the means of evaluation can become routine, as illustrated in Appendix 2.

REFERENCES

1. Axelrod CW (1971) *The Allocation of Computing Resources in Organizations with Semi-Autonomous Users.* Ph.D. Dissertation. Cornell University.
2. Bourne CP, Peterson GD, Lefkowitz B and Ford D (1961) *Requirements, Criteria and Measures of Performance of Information Storage and Retrieval Systems.* Final Report to the N.S.F. on S.R.I. Project 3741.
3. Brenner JR (1965) Toward a value theory of information. In *Economics of Automatic Data Processing* (Ed. Frielink AB) North Holland Publishing Co., Amsterdam.
4. Cassidy RG, Kirby MJL and Raike WM (1971) Efficient distribution of resources through three levels of government. *Mgmt Sci.* **17**(8), B462–B473.
5. Cox DR and Smith WL (1961) *Queues.* Methuen, London.
6. Drucker PF (1974) *Management: Tasks, Responsibilities, Practices.* Harper & Row, New York.
7. Emmons H (1969) One-machine sequencing to minimize certain functions of job tardiness. *Ops Res.* **17**(4), 701–715.
8. Fisher GH (1971) *Cost Considerations in Systems Analysis.* Elsevier, New York.
9. Ghanem SB (1975) Computing center optimization by a pricing-priority policy. *IBM Syst. J.* **14**(3), 272–291.

10. GREENBERGER M (1966) The priority problem and computer time sharing. *Mgmt Sci.* **12**(11), 888–906.
11. HUSKEY HD (1965) Economics of on-line and batch processing computing in a university. In *Economics of Automatic Data Processing* (Ed. FRIELINK AB). North Holland Publishing Co., Amsterdam.
12. LEBON C (1974) *An Opportunity Costing Model for the Evaluation of Scheduling Rules.* Paper presented at the ORSA/TIMS Joint National Meeting, San Juan.
13. MARCHAND M (1968) Priority pricing with application to time-shared computers. In *Proc. AFIPS Fall Joint Computer Conference*, San Francisco.
14. MCNAUGHTON R (1959) Scheduling with deadline and loss functions. *Mgmt Sci.* **6**(1),1–12.
15. MUELLER MW (1960) Time, cost and value factors in information retrieval. In *General Information Manual: Information Retrieval Systems Conference.* Brochure E20-8040, IBM, White Plains.
16. NAOR P (1969) The Regulation of Queue Size by Levying Tolls. *Econometrica* **37**(1), 15–24.
17. NIELSEN N (1968) Flexible pricing: An approach to the allocation of computer resources. In *Proc. AFIPS Fall Joint Computer Conference*, San Francisco.
18. *Proposal to the National Science Foundation Office of Computing Activities for Support of New Facilities in the Information Processing Center.* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts (1968).
19. SHARPE WF (1969) *The Economics of Computers.* Columbia University Press, New York.
20. SINGER N, KANTER H and MOORE A (1968) Prices and the allocation of computer time. In *Proc. AFIPS Fall Joint Computer Conference*, San Francisco.
21. SMIDT S (1968) Flexible pricing of computer services. *Mgmt Sci.* **14**(10), B581–B600.
22. STREETER DN (1972) Cost–benefit evaluation of scientific computing services. *IBM Syst. J.* **11**(3), 219–233.
23. STREETER DN (1975) Productivity of computer-dependent workers. *IBM Syst. J.* **14**(3), 292–305.
24. VICKREY WS (1968) *Economizing and Pricing of Computer Services.* Center for Advanced Study in the Behavioral Sciences.
25. WILLIAMSON OE (1966) Peak-load pricing and optimal capacity under indivisibility constraints. *Am. Econ. Rev.* **66**(4), 810–827.

ADDRESS FOR CORRESPONDENCE: *Dr. C Warren Axelrod, Securities Industry Automation Corporation, 55 Water Street, New York, NY 10041, USA.*

APPENDIX 1

The meaning of value

Value is usually related to benefits [22] or cost savings. Brenner [3] shows, however, that cost minimization does not necessarily lead to net value maximization, since both out-of-pocket expenses and opportunity costs must be considered [8, pp. 24–63, 10]. Sharpe [19, p. 11] defines gross value as the maximum amount someone is willing to pay. Singer *et al.* [20] equate value with the satisfaction derived by the buyer and indicate that such satisfaction must be at least as great as that derived from the money paid. Marchand [13] introduces the concept of social utility, wherein the overall utility of running a job should compensate the disutility its presence generates in others.

Out-of-pocket expenses are usually readily measured, whereas opportunity costs are often less tangible and hence less measureable. Some opportunity costs are so intangible, such as the value of the demand of prospective users [18], that gross estimation is the only approach. The following section surveys the literature on the measurement of the more tractible components of value.

The measurement of value

Mueller [15] suggests three methods for estimating value for computerized information retrieval depending on the cost of retrieval and the time savings. Huskey [11] relates the value

TABLE 1. SYSTEM PERFORMANCE MEASURES FOR VARIOUS MICRO-SEQUENCING RULES FOR LOW AND HIGH ARRIVAL DENSITIES IN TWO MACRO-SEQUENCING ENVIRONMENTS

Macro-sequencing environment	Arrival density	Micro-sequencing rule	No. of jobs run	Total processing (hr)	Average turnaround* (hr)	Variance of turnaround* (hr)	Total value loss** (money units)	Total net value** (money units)	Average net value** (money units)
A. Random arrivals	Low (53 jobs)	FCFS	31	16·76	1·70	1·10	5824	1593	30·05
		SJF	34	16·09	1·39	0·69	5659	1758	33·16
		Priority	26	16·73	1·67	1·03	5101	2316	43·69
		c/t	30	17·46	1·68	1·24	4469	2948	55·62
B. Random arrivals	High (127 jobs)	FCFS	47	22·29	2·04	1·21	15298	1838	14·47
		SJF	55	21·42	1·58	1·06	14991	2145	16·89
		Priority	42	22·79	1·96	1·30	14887	2249	17·71
		c/t	50	23·71	1·81	1·39	13683	3453	27·19
C. Appointment	Low (53 jobs)	FCFS	44	28·08	0·64	0·08	3066	4351	82·10
		SJF	51	29·80	0·58	0·07	1838	5579	105·26
		Priority	44	28·44	0·65	0·08	1777	5640	106·41
		c/t	53	31·96	0·60	0·08	1317	6100	115·09
D. Appointment	High (127 jobs)	FCFS	61	40·27	0·66	0·09	14048	3088	24·31
		SJF	88	38·70	0·44	0·03	9695	7441	58·59
		Priority	70	41·73	0·60	0·06	9118	8018	63·13
		c/t	75	40·79	0·54	0·05	7327	9809	77·23

* For processed jobs.
** For all jobs submitted.

of computer jobs to the type of use, and classifies use according to computer-dependency. Streeter [23] considers both system and user costs to develop priority assignments.

Whereas Huskey and Streeter deal mainly with the classification of users into groups and then estimate value based on the characteristics of the groups, Bourne [2], Cassidy *et al.* [4] and Mueller [15] have attempted to develop explicit relationships between waiting time and value loss. These relationships are exemplified by the value hills of Fig. 3. The more difficult task of developing more complex value hills has not been attempted in the literature, but would be an extension of the above methods.

APPENDIX 2

Simulation parameters

Two macro-sequencing environments were simulated, namely, random arrivals and arrivals by appointment, the former having value hills as in Fig. 3 and the latter as in Fig. 4. For each environment, a low and a high arrival density was used, and within each density category four micro-sequencing rules were applied, namely, FCFS, SJF, priority and c/t.

The duration simulated was 40 hr, and the job processing times ranged from 0.1 to 1.2 hr. The maximum value of a job was between zero and five hundred money units and was related to the inverse of the job priorities, of which there were five. All job characteristics, except for the value hill form, were the same for a particular arrival density for each environment. For low density, 53 jobs were submitted, with 127 for high density.

Simulation results

Table 1 summarizes the results of the simulation runs. Note that in the appointment environment jobs are processed immediately upon arrival, so that queueing time is zero and turnaround time equals processing time.

In general, more jobs are processed for high density than for low density, and the total net value achieved is greater. However, the average net value is less for the high density cases. Also, for random arrivals, the average and variance of turnaround time are greater for higher density, as would be expected for a more congested system.

Within each environment/density category (labelled A, B, C, D in Table 1), the highest number of jobs is run for SJF (except for C), but they yield the lowest total processing. This is due to the SJF rules bias towards running many short jobs. The priority and c/t rules are more favorable when evaluated in value terms since they are based on value considerations. It is interesting to note that the latter are more effective despite being generally less efficient than the SJF rule, if efficiency is measured in terms of processing hours and/or turnaround time.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

A Policy Lapsation Model

VARSHA S VARDE

(*Received October 1975; in revised form February 1976*)

Lapsation of a life insurance policy is discontinuation of premium payment by the policy holder during the period of operation of the policy, due to any reason other than the death of the policy holder. The length of life of a lapsed policy can be defined as the period between the month when the last premium instalment was paid and the month the policy was issued. In India, the acceptance of a proposal for life insurance necessitates administrative processes which, together with the agent's commission and medical charges, cost the Life Insurance Corporation almost the whole of the first year's and a major part of the second year's premium. Early lapses, therefore, pose a major financial problem to the Corporation. In this study, the pattern of early policy lapses was empirically investigated with a view to formulating a model for the lapsation phenomenon. It has been realized that the phenomenon is not amenable to any simple statistical model due to an inherent stratification in the population of lapsed policies.

Vol. 4
1976

1. INTRODUCTION

THIS PAPER presents a model that is developed for explaining policy lapsation phenomenon in Life Insurance Corporation (LIC) of India. The policy lapsation phenomenon in LIC of India goes almost unnoticed by the researchers. Withdrawal experiences in the U.K. have been studied by Scobbie and Patrick [4], giving special emphasis on the rates of withdrawal and the factors influencing withdrawal rates. The present paper considers the policy lapsation phenomenon in India from the statistical view point and presents a statistical model to explain it. The objective of the project was to understand and study the pattern of policy lapses in the LIC of India. Initially, we analysed the policy lapsation data with a view to finding out whether one of the existing statistical distributions can explain the behaviour of policy lapses. None of the well-known statistical models such as Poisson, binomial, negative-binomial fitted to the data on policy lapsation satisfactorily. A new model was therefore conceptually developed and based on the rationale behind the policy lapsation phenomenon. A composite model emerged out of this process, which could be termed as a Binomial–Poisson model. This distribution has not been explored much in the theory of statistics; a passing reference has been made by Khatri and Patel [2], and Katti [1], while considering the whole class of compound distributions. After

developing its statistical properties and estimation procedures, this distribution was tried on the data on policy lapses obtained from the LIC. It offered an excellent fit.

Failure or lapsation of a policy is defined as the discontinuation of premium payment by the policy holder during the period of operation of the policy due to any reason other than the death of the policy holder. Length of life of a failed policy is defined as the time elapsed between the payment of the last premium and the issue of the policy.

2. DEVELOPMENT OF THE MODEL

Let Y be the length of life of a lapsed life insurance policy. Policies fail at discrete moments and so Y takes only discrete values. The situation $Y = k$ is equivalent to receiving exactly k premium instalments on the policy concerned, subsequent to the first premium which has to accompany the application.

Actual data on the frequency of policy failures distinctly reveals that the probability of zero life is very high and the probabilities of subsequent life lengths keep on dropping rapidly. The situation is thus compatible with those characterized by the Poisson distribution. The random variable Y under present consideration can be, therefore, assumed to follow a Poisson distribution.

Vol.
197

However, it is worth noting that not all policies behave according to a single pattern. Policies sold by commission agents towards the end of the year merely to fulfil the business targets generally lack 'will' on the part of the customers and the mean life of these policies is found to be far shorter than that of the policies genuinely required by the customers. (See [3]). Similarly, a policy of a huge sum assured held by a person with insufficient disposable surplus is likely to lapse much earlier than the rest of the policies. Life insurance policies can thus be classified into different groups in such a way that

(i) the policies within a group possess almost identical mean lives, and
(ii) the mean life varies significantly from group to group.

Consider a set of basic factors necessary for survival of life insurance policies. Examples of such factors are, genuineness on the part of the customer (applicant) and the agent, sound financial condition of the customer relative to the commitment, effective organisation for reminding and collecting premium instalments, etc. These factors may be termed stimulative factors or stimulants. Naturally, the number of such basic stimulants would not be very large. Absence of all of them would result in the immediate lapsation of the policy, as they are the necessary conditions for continuation of the policy. Presence of a large number of stimulants would increase the probability of its staying on which will be reflected in the improved mean life. Let n be the number of possible independent stimulants in respect of a life insurance policy. Let the probability of presence

of the stimulant i be denoted by P_i and let the contribution made by this stimulant to the mean life of the concerned policy be denoted by $\lambda_i, i = 1,2,\ldots n$. Since the stimulants are independent, the presence of any one of them does not in any way affect the probability of the presence of any other stimulant. If S is the set of stimulants, out of these n stimulants, actually underlying a policy, then the probability function of S is

$$Pr\,(S) = \prod_{i \in S} P_i \prod_{j \notin S} (1 - P_j). \tag{1}$$

It is obvious that the probability that S is a null set is $\prod_{i=1}^{n} (1 - P_i)$.

Let a set consisting of k stimulants ($k = 1,2,\ldots n$) be called a stimulant set of order k and be denoted by S_k. There will exist $\binom{n}{k}$ specific sets of order k. If S_k is a set of stimulants underlying a policy then the mean life of that policy is a function of the contributions to the mean life made by the stimulants present in S_k. Let this function be denoted by $f(S_k)$.

Further the conditional function of the life length of a policy under the influence of the stimulant set S_k is Poisson with mean $f(S_k)$. In view of (1) the unconditional probability function of Y is

$$Pr\,(Y = y) = \sum_{k=1}^{n} \sum_{\binom{n}{k}} e^{-f(S_k)} \frac{[f(S_k)]^y}{y!} \prod_{i \in S_k} P_i \prod_{j \notin S_k} (1 - P_j)$$

where Σ denoted the sum over the $\binom{n}{k}$ possible stimulant sets of order k and $\binom{n}{k}$

$$Pr\,(Y = 0) = \sum_{k=1}^{n} \sum_{\binom{n}{k}} e^{-f(S_k)} \prod_{i \in S_k} P_i \prod_{j \notin S_k} (1 - P_j) + \prod_{i=1}^{n} (1 - P_i).$$

This is the model explaining the failure phenomenon in respect of life insurance policies. It is, however, rather complex to employ in practice for drawing the necessary inferences such as those based upon estimates of parameters given a set of data. In order to make the above model amenable to further mathematical treatment without compromising much on reality, let us make the following assumptions.

1. *Additivity of contributions.* Let the joint contribution to the policy mean life by individual stimulants underlying the policy be the sum of the individual contribution. That is if $i_1, i_2, \ldots i_k$ are stimulants in a specific S_k then

$$f(S_k) = \lambda_{i_1} + \lambda_{i_2} + \ldots + \lambda_{i_k}.$$

This assumption has been made in view of the stochastic independence of the stimulants and an observation made earlier that the presence of a larger number of stimulants would increase the mean life of the policy.

2. *Equi-contributivity of stimulants.* Let the contributions independently made by the stimulants be identical, that is $\lambda_1 = \lambda_2 = \ldots = \lambda_n = \lambda$,

Vol. 4
1976

In other words, all stimulants would be equally important individually in improving the mean life of the policy.

3. *Equi-probable occurrence of stimulants.* Let the individual probabilities of the presence of stimulants in respect of a policy be indentical, that is $P_1 = P_2 = \ldots P_n = p$.

If the already defined stimulants do not possess the last two properties within an acceptable range, splitting and grouping of them can be done in such a way that the finally defined set of stimulants possesses these properties. It may be noted that the process of splitting and grouping does not necessarily retain the basic property of stochastic independence. Therefore, it will be necessary to check whether the newly defined stimulant groups possess this property. The process of splitting and grouping thus may require some iterations to arrive at a satisfactory grouping of stimulants. It still leaves the possibility of not finding stimulant groups possessing the required properties within the permissible margin of deviation. Further simplification of the model in such situations is not possible.

Under the three assumptions stated above, if we define X as the number of stimulants underlying a policy, then X would follow a binomial distribution with parameters n and p, where n is the maximum possible number of stimulants (that is, stimulant-groups) satisfying the above mentioned properties. Note that n would then be a known integer. Presence of exactly x stimulants out of these n would make the mean life of the concerned policy equal to λx for $x = 0, 1, 2\ldots, n$. We thus have the following formulation of the model:

Let Y be the length of life of a life insurance policy and X be the number of stimulants underlying the policy. Then X would follow a binomial distribution with parameters n and p where n is the maximum possible number of stimulants and p is the probability of the presence of a stimulant. If λ is the contribution made to the mean life of the policy by the presence of a stimulant, then the probability function of Y when exactly x stimulants are present is Poisson with parameter (mean life) λx. In mathematical terms

$$Pr(Y = y/X = x) = \frac{(\lambda x)^y}{y!} \cdot e^{-\lambda x} \tag{2}$$

for $y = 0, 1, 2, \ldots$ and

$$Pr(X = x) = \binom{n}{x} p^x q^{n-x} \tag{3}$$

for $q = 1 - p$ and n known.

The unconditional probability function of Y, the length of policy life in which we are interested is

$$Pr(Y = y) = \sum_{x=0}^{n} Pr(Y = y/X = x) \,.\, Pr(X = x)$$

or

$$Pr(Y = y) = \frac{\lambda^y}{y!} \sum_{x=0}^{n} x^y \binom{n}{x} (pe^{-\lambda})^x q^{n-x} \tag{4}$$

3. EMPIRICAL TESTING OF THE BINOMIAL–POISSON MODEL

In order to test the conceptually developed model empirically, data were obtained on lapsed policies out of the policies issued by the Bombay Divisional Office of the Life Insurance Corporation of India during November, 1970.

Using the estimation procedure (developed in [5]) the maximum likelihood estimates of λ and p for $n = 3$ were obtained and were found to be

$$\hat{\lambda} = 1{\cdot}6872, \qquad \hat{p} = 0{\cdot}2065.$$

The reasons for selecting $n = 3$ were as follows: as described in Section 3 the parameter n is interpreted as the maximum possible number of stimulants satisfying the three conditions, namely, additivity of contribution, equi-contributivity of stimulants and equi-probable occurrence of stimulants. To determine n this way, a complete knowledge of factors influencing operation of life insurance policies is required (see [3]) On the basis of this information it was possible to form independent groups of stimulants satisfying the condition of additivity.

The following three groups of stimulants were found to satisfy the properties of independence and additivity.

1. Genuineness both on the part of the agent and the customer (generally called absence of 'benami' (fictitious) nature in the transaction).
2. Financial ability of the policy holder to pay premium instalments regularly.
3. Efficient organizational arrangements for reminding the policy holder and for collecting the premium instalment. (This reflects on the role of the Corporation as well as the agent.)

The conditions of equiprobability and equicontributivity cannot be, however, demonstrated in view of the possible non-homogenity of raw data by plan of assurance, sum assured, operating zone etc. However, the model might still be expected to give a reasonable representation in spite of the variety of classes within the observed policies.

Using these estimated values of λ and p, the expected frequencies for various values of y using (4) were computed. Table 1 presents results of the computation along with the actual observations.

TABLE 1. MAXIMUM LIKELIHOOD FITTING OF BINOMIAL–POISSON DISTRIBUTION

Life length	Observed frequency	Expected frequency
0	136	135
1	29	31
2	28	29
3	20	19
4	11	10
5	6	5
$\geqslant 6$	5	6
Total	235	235

The last frequency is adjusted so as to make the total of expected frequencies equal to that of the observed. The chi-square goodness of fit works out to be 0·6902 when the 5 per cent table value is 11·070. The binomial–Poisson model, thus, fits well to the actual data obtained from the Life Insurance Corporation of India. It may be noted that this excellent fit is obtained even for a very small value of n and thus a larger value of n will certainly provide a satisfactory fit between the observed and expected frequencies. More extensive data are required before the size of n could be limited to a small integer.

CONCLUDING REMARKS

This paper has developed the binomial-Poisson distribution as a quantitative model for explaining the phenomenon of failure of life insurance policies. Its significance lies in enabling us to make useful predictions. For example, given the long-term annual performance plans of the LIC, we can predict the number of policies which would lapse before they start contributing positively to the Corporation's profit and loss account.

The binomial–Poisson model offers an excellent fit to Indian data on policy lapses. But, it does not provide a satisfactory fit to the crude rates of withdrawal given in Table 1 of the paper by Scobbie and Patrick [4]. The disagreement between the model and that data is mainly around the zero-life withdrawals. In India, half the withdrawals occur in the first year itself, while the British experience is not so grave. Roughly 20% withdrawals in the UK occur during the first year. There is thus a difference in the pattern of lapsation between the two countries. It may be recalled that the use of the Poisson distribution in Section 3 of the paper was prompted by the observation that the probability of zero life was very high. Efforts to develop a statistical model to explain the British experience should therefore take cognizance of this fact and select a base distribution according to the lower zero life probability and gradual tapering off of the subsequent probabilities.

One of the decisions to be made before employing this model is whether a single model would suffice or the heterogeneity present in the group of policies being studied demands stratification. In the latter situation it would be necessary to evolve a binomial–Poisson model for each stratum. Heterogeneity in the lapsation phenomenon may occur on account of the following factors:

(a) *Plan of assurance.* The Corporation has a variety of assurance schemes such as whole life assurance, endowment assurance, marriage endowment and educational annuities, etc. If the policies under consideration are of a variety of types, it would be useful to work out models for policy-groups of each type and test whether the plan of assurance really affects the policy behaviour. In case it does, the quality of forecasts could be improved by obtaining them separately for each category.

(b) *Sum assured.* Experience of some of the insurance agents initiates the hypothesis that the bigger policies are more prone to lapsation than the smaller ones. Testing this hypothesis has policy level implications. Decisions regarding incentive schemes for policies of different magnitudes could be taken accordingly.

(c) *Mode of payment.* This factor may also have some effect on the behaviour of policies. Findings throwing light on whether quarterly, half yearly or yearly mode of payment for a particular section of the policy holders would help develop a stable relationship with the Corporation would be of immense assistance in the business promotion efforts.

(d) *Operating zones.* Different zones operate in different environment possibly resulting in different failure rates. While the Corporation experienced an average of 18–24% lapses during the last decade, certain places in the Northern India experienced a lapse percentage as high as 80%. This situation necessitates zonal or even divisional (LIC has 38 divisions grouped in 5 zones) predictions of failures which would be of considerable importance in modifying the management style and organisation structure of individual operating units.

Vol. 4
1976

ACKNOWLEDGEMENT

The author is indebted to Dr. YS Sathe, Head of the Department of Statistics, University of Bombay, for his valuable guidance during the course of this work. The author also wishes to express her gratitude to the referees for their helpful suggestions which led to improvements in an earlier draft of this paper.

REFERENCES

1. KATTI SK (1966) Inter-relations among generalized distributions and their components *Biometrics* **22**, 44–52.
2. KHATRI CG and PATEL IR (1961) Three classes of univariate discrete distributions. *Biometrics* **17**, 117–24.
3. Report of Committee of Enquiry into the Expenses of Life Insurance Corporation of India (1969) Parts 1 and 2, Manager of Publications, Govt of India Press, Delhi.

4. SCOBBIE A and PATRICK FD (1969) Some aspects of withdrawals in ordinary life business. *Transactions of the Faculty of Actuaries*, **231**, 53–93.
5. VARDE VS (1974) *Some Problems of Estimation in Discrete Distributions and In Accelerated Life Testing*. Ph.D Dissertation (Unpublished), University of Bombay.

ADDRESS FOR CORRESPONDENCE: *Dr. VS Varde, National Institute of Bank Management, 85 Nepean Sea Road, Bombay – 400 006, India*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Memoranda

Inventory Decisions in Product-Mix Models

A PRODUCT-mix model is one in which scarce resource inputs are transformed into an optimal product-mix of outputs. A common vehicle for solving such models is linear programming. In many instances, various combinations of resource inputs (right hand side coefficients) and/or product outputs (optimal values of the decision variables) are subject to inventory decisions. However, economic order quantities for the inputs and economic production quantities for the outputs are typically determined outside the context of the product-mix model. Separate consideration of interrelated decisions can lead to suboptimal results. The purpose of this note is to formulate several models which integrate product-mix and inventory decisions and to apply an existing algorithm to their solution.

A retail sales-mix model might be stated as:

$$\text{Maximize } f(x) = \Sigma\, c_j x_j \tag{1}$$

$$\text{subject to } \Sigma\, a_{ij} x_j = b_i \; (i = 1, \ldots, m) \tag{2}$$

where Σ stands for $\sum_{j=1}^{n}$ and where the decision variables $x_j \geq 0$ (j=1, ..., n) represent the number of units of product j to sell over a given time period. The model coefficients, the c_j's, b_i's and a_{ij}'s, represent the per unit profit, the initial amount of resource i available, and the amount of resource i consumed per unit of product j, respectively.

Once the optimal values of the decision variables in the linear programming model, the x_j^*'s, have been determined, the retailer is faced with the decision of how often and in what quantities to order each of the products from the manufacturer. Using standard economic order quantity analysis, the inventory cost for ordering Q_i is given by:

$$IC_j = \frac{K_j x_j^*}{Q_j} + \frac{h_j Q_j}{2}, \tag{3}$$

where K_j and h_j are the order and holding costs, respectively; the economic order quantity which minimizes IC_j is

$$Q_j = \sqrt{\left(\frac{2.K_j.x_j^*}{h_j}\right)}. \tag{4}$$

so that

$$IC_j = \theta_j.x_j^{*1/2} \text{ [where } \theta_j = \sqrt{(2.K_j.h_j)}\text{]}. \tag{5}$$

Incorporating the nonlinear inventory costs (5) into the objective function (2) results in:

$$\text{Maximize } f(x) = \Sigma\,(c_j x_j - \theta_j x_j^{1/2}) \tag{6}$$

subject to (2). This model is very similar to one proposed by Raiborn and Harris [2].

Next, consider the problem faced by a manufacturer, where some combination of resource inputs, b_i's, might be subject to inventory decisions, as were the outputs of the retail model. It is possible that by treating these inputs as decision variables and integrating their nonlinear inventory costs into the product-mix objective function, he may be able to achieve a greater total profit than by considering product-mix and inventory decisions separately. Without loss of generality, let us assume that we wish to integrate inventory decisions concerning the first r resource inputs, $0 \leq r \leq m$, into the product-mix model and adopt the notation

$$\left.\begin{aligned} I_1 &= \{i | i \leq r, i = 1, \ldots, r\} \\ I_2 &= \{i | i > r, i = r+1, \ldots, m\} \end{aligned}\right\} \tag{7}$$

and replace b_i with a new decision variable x_{n+i} for $i \in I_1$. The corresponding integrated model would be:

$$\text{Maximize } f(x) = \Sigma\, c_j x_j - \sum_{i \in I_1} \theta_{n+i} x_{n+i}^{1/2} \tag{8}$$

$$\left.\begin{aligned} \text{subject to } \Sigma\, a_{ij} x_j - x_{n+i} &= 0 \quad i \in I_1 \\ \Sigma\, a_{ij} x_j &= b_i \quad i \in I_2 \end{aligned}\right\} \tag{9}$$

and again $x_j \geq 0$.

Vol. 4
1976

When a combination of both inputs and outputs is subject to inventory decisions, the following integrated model might be appropriate:

$$\text{Maximize } f(x) = \sum [c_j x_j - \theta_j x_j^{1/2}] - \sum_{i \epsilon I_1} \theta_{n+i} x_{n+i}^{1/2} \quad (10)$$

subject to (9).

Since the objective functions of (6), (8) and (10) are convex, global maximum solutions occur at extreme points of their respective solution spaces. However, there may be local maxima different from the global maximum. Inventory setup and order costs can be thought of as fixed charges. If every decision variable is subject to an inventory setup or order cost, then every extreme point of the solution space is a local maximum. As would be expected, this fact greatly complicates the problem of determining a global solution.

Soland [3] has proposed a finite branch and bound algorithm for the global maximization of a separable convex function over linear polyhedra. It is a simplified version of an earlier algorithm by Falk and Soland [1] and can be applied to problems described here.

An advantage of an integrated model is that it represents a closer, more realistic approximation to the actual problem situation. A potential disadvantage is that the effort required for its solution tends to be greater than the sum of the efforts required to solve separate product–mix and inventory models.

REFERENCES

1. Falk JE and Soland RM (1969) An algorithm for separable nonconvex programming problems. *Mgmt Sci.* **15** (9).
2. Raiborn MH and Harris WT (1974) Integration of inventory and product sales–mix models. *Decis. Sci.* **5** (4).
3. Soland RM (1974) Optimal facility location with concave costs. *Ops. Res.* **22** (2).

GR Reeves
(*February 1976*)

Assistant Professor
School of Business Administration
Miami University
Oxford, Ohio 45056
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Personnel Development and Evaluation

THE EFFECTIVENESS of any organization is dependent on the quality of its personnel. The right people must be originally selected into the organization, motivated to work; and sound personnel promotion and training decisions must be made in filling nonentry level jobs. An effective personnel performance evaluation system is a crucial cornerstone in this process, as it provides the data needed for most of the required administrative decisions. This system plays a key role in motivating people to utilize their abilities in pursuing the organization's goals.

The research reported here dealt with developing an improved performance evaluation system for civilians in two US Navy organizations employing both uniformed and civilian personnel. The research project had four major parts:

Group appraisals

The management development program involved civilian professionals at the two commands in group performance appraisal sessions. The professional's work performance was appraised by his supervisor, the supervisor's superior, and a member of the research team. The appraisals conducted by these groups were focused on, and limited to, *intra*-appraisee considerations. That is, the appraisal committee considered the individual in terms of his greatest strengths and his least strong work performances, but did not compare the appraisee with other individuals. Recommendations for the appraisee, for the appraisee's supervisor, and for the organization were made so that the appraisee (a 'human asset') could grow in worth to himself and to the organization.

The supervisor then discussed the appraisal with the appraisee, stating that this is "how others see and interpret you," and that "here are our thoughts on how you might further develop and utilize your talents."

The development of supervisory skills in these feedback behaviors appears to be a crucial requirement if face-to-face dialogues between supervisors and subordinates are to become common and meaningful [1].

Goal setting

As another part of the research, a framework for results-oriented management was established using management by objectives techniques. Efforts were directed toward developing organizational objectives, and then deriving the objectives for units and individuals so all of these objectives would be congruent. A goal setting manual and a work performance folder were developed for the organization. These two instruments can be used to formulate goals, monitor and control performance, and to appraise work performance at the end of the year [2].

Performance rating scales

Another portion of the research project focused on developing performance evaluation rating scales providing data for administrative decision making (e.g. promotions, pay changes, etc.) concerning civilian professionals in the two organizations. Four steps were followed in developing these rating scales: (1) A group of employees (3–6) familiar with the job listed the most relevant aspects of performance for the specific job. (2) The group then generated 'specific' behavioral examples they had observed that demonstrated high and low performance on each performance aspect. (3) At a later time, these behavioral incidents were presented to the individuals in the group, who assigned them to the rating scale (aspect) and rating scale level (low to high on a 5-point scale) that they thought appropriate. (4) Incidents that were not by consensus assigned to the same location (both rating scale and level) were eliminated.

This procedure yielded rating scales that are relevant to the specific job and that are 'anchored' by specific behavioral incidents representing on the scales the various levels of job performance [3].

Career paths

A task inventory asking civilian professionals to list the degree to which they were involved in various activities was administered at the two organizations. Data

Vol. 4
1976

from these inventories were subjected to canonical correlation and cluster analysis methods in an attempt to find linkages (career paths) among the jobs in the organizational hierarchies. The results indicated that neither the analytical methods nor the task inventory was wholly adequate [4,5].

CONCLUSION

The main conclusion of this study was that the sources of deficiencies in the development or evaluation of personnel lie not in flaws of the methodology, or technology of performance evaluation, or even the rating forms available, but emanate from the absence of responsibility on the part of the line manager to accept personnel development and performance appraisal as part of 'his job.' There were some exceptions, but in general very little attention was given to evaluation of the work performance, or to the training and development of civilian professionals. The typical line manager thought of the training or development and, sometimes, the disciplining of his subordinates, as being the responsibility of the 'Personnel Department.' This was the prevailing concept, despite contrary statements by the Personnel Director and direct conflict with current official directives.

REFERENCES

1. GITHENS W, ELSTER R, MUSGRAVE G and CREIGHTON J (1973) *Design of an Operational Management Development Manual.* Naval Postgraduate School Technical Report No. NPS-55Gh73062, 1974.
2. MUSGRAVE G, ELSTER R, CREIGHTON J and GITHENS W (1973) *Design of an Operational Management by Objectives Manual.* Naval Postgraduate School Technical Report No. NPS-55Mf73061, 1974.
3. GITHENS W, ELSTER R, MUSGRAVE G and CREIGHTON J (1973) *Design of an Operational Rating Manual.* Naval Postgraduate School Technical Report No. NPS-55Gh73063, 1974.
4. ELSTER R, READ R, GITHENS W, MUSGRAVE G and CREIGHTON J. (1973) *Design of Operational Career Ladders.* Naval Postgraduate School Technical Report No. NPS-55Ea73061, 1974.
5. READ R, ELSTER R, MUSGRAVE G, CREIGHTON J and GITHENS W (1973) *Statistical Analysis of Personnel Data Using Factor Scoring, Cluster Analysis, and Multi-Dimensional Scaling.* Naval Postgraduate School Technical Report No. NPS-55Re73061, 1974.

WH Githens
RS Elster
***GL Musgrove**
JW Creighton
(*February 1975*)

**Hoover Institution*
Stanford University
Stanford, Calif 94305
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

The Value Index Concept for Demand Forecasting

FORECASTS of demand are required as part of the product development cycle. Traditional methodologies may fail if the new product is significantly different from previous ones. There is, however, commonality of attributes of products which satisfy the same need in the marketplace. In this note a method for translating product history into a need satisfaction value index is described. Statistical and analytic techniques can then be used to forecast demand for this value index which in turn can be interpreted as a product forecast.

Traditional methods used for new product demand forecasting include time series analysis and market research. Time series analysis can be used when a new product is similar to products already on the market for which time series data are available. Market research studies are used when historical data on similar products are unavailable. The method described herein can be used in lieu of a market research study providing market data are available on dissimilar products which have, however, satisfied similar needs in the past. Two examples are products in the transportation and computer industries.

In brief, the method defines value received as a function of product attributes. The value index so defined is a surrogate commodity measure. Parameters of a value index function are determined from historical data. Once the function is defined, a time series of value indexes can be determined and projected. The forecasted value index can then be used as a commodity measure in econometric models to determine demand as a function of product attributes. A value index can be defined in terms of either utility theory or production theory. The concept is illustrated here by use of production theory.

The theory of production [3] describes the output of a firm as a function of the factors of production. We equate value (or value index) with output, user with firm, and product attributes with factors of production. Let x_j be a measure of the *j-th* factor for $j = 1, ,... n$. For a given technology and level of expenditure, C, the production function relates value, $V^*(C)$, to the level of the factors used. Associated with a given technology are unit prices, c_j, for each factor. The key assumption is: *Every user attempts to maximize his value for a given level of expenditure* (or minimize his cost for a required value). Differences in value received for any level of expenditure occur due to differences in perception, processes, entrepreneurship, and preferences. The value for any given user, $V(C)$, is always less than or equal to the best value, $V^*(C)$, for the available technology. The set of all possible values considering different expenditures is called the production set.

For any hypothesized class of functions, there is a vector, $\mathbf{w}$, of parameters which distinguish members of the class. Of the several well known functions used in production theory, a variation of the Cobb–Douglas [1] was selected because it requires the fewest number of parameters. This function is given by

$$V = w_0 \prod_1^n x_j^{w_j} \qquad (1)$$

where $w_j > 0$, $\sum_1^n w_j = r$, and w_0 is an arbitrary scale constant. The best value that can be obtained for a given technology and expenditure in (1) subject to $\sum_1^n c_j x_j \leq C$ and $x_j \geq 1$ is given by the solution $x_i = Cw_j/rc_j$ and $V^*(C) = (C/r)^r \prod_1^n (w_j/c_j)^{w_j}$.

The most representative value function in this class can now be defined as the function which minimizes the difference $V^*(C) - V(C)$ for all C. Each observation, i, for $i=1, ..., m$ consist of a level of expenditure, C_i, and factors, x_{ij}. The user perceived value is unknown, but we assume the user tried to obtain the most value. The problem can now be stated formally as: Find w_j for which minimize

$$\sum_{i=1}^m (C_i/r)^r \prod_1^n (w_j/c_j)^{w_j} - \prod_1 x_{ij}^{w_j},$$

Vol. 4
1976

where $C_l = \sum_{1}^{m} c_j x_j$. This is an unconstrained minimization problem which can be solved by a variety of nonlinear optimization methods [2]. The vector **w** of parameters obtained by solving this problem along with an arbitrary scaling constant, w_0, define the value index (1).

A number of observations on the characteristics of this function can be made. The w_j are the factor elasticities of V; the variable r is the returns to scale of V (assuming *a priori* knowledge of r); the marginal physical productivity of x_j is Vw_j/x_j and declines with increasing x_j if $w_j < 1$. The marginal rate of substitution of x_j and x_k is $w_j x_k / w_k x_j$, and so the elasticity of substitution is one.

The production theory interpretation of the value index with regard to two types of technological change is important. One is a change, in time, to the unit prices of factors. Different prices for a factor can thus be used to determine a single set of **w**. A technological change which results in the availability of a new product attribute, on the other hand, may affect the choice of factors x_j. It is clear that this type of technological change must result in a new definition of **w**.

Many possibilities exist for the use of this concept for different types of products. Different function classes for the value index can also be defined. The essential concepts common to all these possibilities remain the same. The value index is defined in terms of measurable product attributes, and parameters are determined in such a way as to reflect actual user perception of the contribution of those product attributes to overall product desirability.

REFERENCES

1. Cobb CW and Douglas PH (1928) A theory of production. *Am. Econ. Rev.* **18,** (Suppl.), 139–165.
2. Fletcher R and Powell MJD (1963) A rapidly convergent descent method for minimization. *Computer J.* **6,** 163–168.
3. Walters AA Production and cost functions: an econometric survey. *Econometrica* **31** (1–2), 1–66.

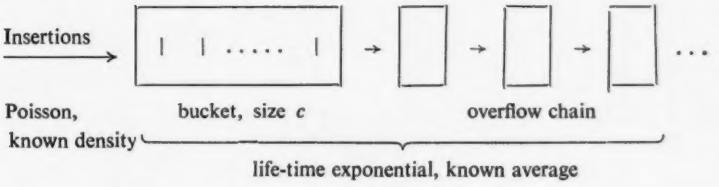
Gerald L Hefley
(*December 1975*)

37 Benson Drive
Danbury, Ct 06810
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Overflow Problems in File Organization

WHEN using indirect addressing of records in random access files, certain stochastic problems arise in connection with 'overflow'. It is not generally appreciated that alike problems occur in telephony and have been studied ever since 1918. Let us consider the following case. A large file of fluctuating size is split up into groups by an effective form of 'hashing' of the keys, the size of the groups also varying by random insertion and deletion of records. Per group insertions are assumed to occur according to a Poisson process with known density. The life-time of the records is supposed to be distributed exponentially with known average. In connection with each group there is a 'bucket' of size c, in which records of the group have to be placed in first instance. When the bucket is full, new records are placed in a common overflow area, forming a chain for each group. 'Holes' in the bucket, caused by deletions, are *not* filled by transfer from the chain, however. The model per group is the following:



Stationarity is supposed to apply. The following quantities are of interest:

(i) the average loads per bucket and chain;

(ii) the average number of searched records in the overflow-chain needed to reach a record.

Let $\mathbf{r}$ and $\mathbf{s}$ be the momentary values of occupations in the bucket and in the chain, respectively. The average loads may be obtained from the expectations of $\mathbf{r}$ and $\mathbf{s}$. They have been studied by Olson [1]. The result is Erlang's B formula (1918), [2], to be found in many textbooks on queueing, e.g. [3, 4, 5]. This result is independent of the life-time distribution [6, 7, 8], *cf.* also [5].

Van der Pool [9] studied the average number of searched overflow records by deriving a set of linear equations for the joint distribution function of $\mathbf{r}$ and $\mathbf{s}$, truncating this set and solving it numerically. It may be observed that the distribution of $\mathbf{s}$, conditional on $\mathbf{r} = c$, is sufficient in this respect. Explicit expressions for this conditional distribution are known [10, 11, 12]. Those publications are hardly accessible, but summaries can be found in [3, 4, 5]. Even the first and second moments of $\mathbf{s}$, conditional on $\mathbf{r} = c$, are enough. There are simple expressions for these moments [13, 14, 5].

In [13, 14] an ingenious method (called the "Equivalent Random Method" by Wilkinson) has been described for dealing with more complicated overflow problems in an approximate though very accurate way. It seems to be promising for overflow situations connected with 'open addressing', 'multiple hashing', etc.

Vol. 4
1976

REFERENCES

1. OLSON CA (1969) *Proc. of the 24th Conf. of ACM*, New York, 539.
2. ERLANG AK (1918) *P.O. Elect. Engrs' J.* **10,** 189.
3. SYSKI R (1960) *Introduction to Congestion Theory in Telephone Systems.* Oliver & Boyd, London.
4. COOPER RB (1972) *Introduction to Queueing Theory.* MacMillan, New York.
5. KOSTEN L (1973) *Stochastic Theory of Service Systems.* Pergamon Press, Oxford.
6. PALM C (1938) *Ericsson Tech.* **4,** 39–58.
7. KOSTEN L (1948) *PTT-Bedr.* **2,** 42.
8. SEVASTJANOV BA (1957) (Russian, Engl. summary) *Theory Probab. Applic.* **2,** 106.
9. POOL JA van der (1973) *IBM Jl Res. Dev.* **17,** 27.
10. KOSTEN L (1937) *Elekt. Nachr. Tech.* **14,** 5.
11. KOSTEN L (1937) *Annls Post. Télégr. Téléph.* **26,** 1002.
12. KOSTEN L (1942) *Over Blokkeerings- en Wachttijdproblemen.* Doct. diss. Delft.
13. GILTAY J (1953) *Ingenieur* **65,** E 107.
14. WILKINSON RI (1956) *Bell. Syst. Tech. J.* **35,** 421.

L Kosten

(*February 1976*)

Afdeling der Algemene Wetenschappen
Onderafdeling der Wiskunde
Technishe Hogeschool Delft
Delft-8
Julianalaan 132
Netherlands

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

The Simple Plant Location Problem

THE *simple plant location problem* was formulated by Balinski [1] as follows:

Minimize

$$z = \sum_{i=1}^{m} f_i y_i + \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \quad (1)$$

subject to

$$\sum_{i=1}^{n} x_{ij} = 1 \quad j = 1, 2, ..., n, \quad (2)$$

$$0 \leqslant x_{ij} \leqslant y_i \leqslant 1 \quad i = 1, 2, ..., m; \quad j = 1, 2, ..., n, \quad (3)$$

and

$$y_i \in \{0, 1\} \quad i = 1, 2, ..., n. \quad (4)$$

In this note, we show how the study of upper bounds on supply costs in optimal solutions makes it possible to obtain tighter bounds than those previously proposed as well as improved direct and conditional tests. The results are illustrated by the resolution of an example due to Spielberg [4], in which branching does not occur when the new tests are used.

BOUNDS

At a current iteration of the resolution of (1)–(4) by a branch-and-bound algorithm, some variables y_i will be fixed at 1, some will be fixed at 0 and some will be free; let $K_0 = \{i | y_i = 0\}$, $K_1 = \{i | y_i = 1\}$ and $K_2 = \{1, 2, ..., m\}/\{K_0 \cup K_1\}$, $i | i \notin K_0 \cup K$ denote the sets of indices of these variables. The current subproblem may be written:

Minimize

$$z' = \sum_{i \in K_1} f_i + \sum_{i \in K_2} f_i y_i + \sum_{i \in K_1 \cup K_2} \times \sum_{j=1}^{n} c_{ij} x_{ij} \quad (5)$$

subject to

$$\sum_{i \in K_1 \cup K_2} x_{ij} = 1 \quad j = 1, 2, ..., n, \quad (6)$$

$$0 \leqslant x_{ij} \leqslant 1 \quad i \in K_1; j = 1, 2, ..., n, \quad (7)$$

$$0 \leqslant x_{ij} \leqslant y_i \leqslant 1 \quad i \in K_2; j = 1, 2, ..., n \quad (8)$$

and

$$y_i \in \{0, 1\} \quad i \in K_2. \quad (9)$$

Let c_{ij} denote the smallest cost for satisfying the demand of customer j from an already set-up plant:

$$c_{ij} = \text{Min}\,\{c_{ij} | i \in K_1\}; \quad (10)$$

let c_{ij} and c_{sj} denote the smallest and second smallest costs for satisfying the demand of customer j from an already set-up plant or from a plant to be set up:

$$c_{rj} = \text{Min}\,\{c_{ij} | i \in K_1 \cup K_2\} \quad (11)$$

$$c_{sj} = \text{Min}\,\{c_{ij} | i \in K_1 \cup K_2, i \neq r\}. \quad (12)$$

Efroymson and Ray [2] have noted that if

$$g_1(i) = \sum_j \max(c_{sj} - c_{ij}, 0) - f_i \quad (13)$$

is non-negative, $y_i = 1$ in at least one optimal solution of (5)–(9). Let us assume all variables such that $g_1(i) > 1$ have been fixed at 1.

Theorem 1. In any optimal solution of (5)–(9) the supply cost of customer k cannot exceed

$$c^*_{ik} = \min(c_{ik}, c_{vk}) \quad (14)$$

where

$$c_{vk} = \text{Min}\,\{f_i + c_{ik} - \sum_{j=1, j \neq k}^{n} \max(c_{sj} - c_{ij}, 0) | i \in K_2\}. \quad (15)$$

Proof. If $c_{vk} \geq c_{ik}$ the result is obvious. If $c_{vk} < c_{ik}$, let $Y^* = (y_i^*)$ denote the value of Y in an optimal solution of (5)–(9) and c_k^* the supply cost of customer k in that solution.

Vol. 4
1976

TABLE 1.

f_i	3	c_{ij}	2	5	2	12	5	9	6	4	3	4
	3		5	9	6	7	8	5	6	3	7	12
	4		7	10	6	6	3	11	10	7	5	10
	5		4	6	8	11	3	2	12	7	8	4
	5		8	10	4	7	9	8	10	8	3	8
	6		5	9	6	5	11	8	13	12	6	9
	7		7	4	7	3	9	6	2	7	5	13
	8		4	9	4	13	8	10	4	10	9	10
	8		5	12	9	14	12	2	8	8	6	9
	9		9	11	2	9	7	12	10	3	7	7

TABLE 2.

i or j	1	2	3	4	5	6	7	8	9	10
c_{rj}	2	4	2	3	3	2	2	3	3	4
c_{sj}	4	5	2	5	3	2	4	3	3	4
$g_1(i)$	−1	−3	−4	−5	−5	−6	−2	−8	−8	−9
$c^*{}_{tj} = c_{vj}$	5	6	3	7	6	7	6	5	4	5
$g^*{}_2(i)$	6	1	0	5	−4	−4	4	−5	−3	−6
y_i	—	—	—	—	0	0	—	0	0	5
c_{rj}	2	4	2	3	3	2	2	3	3	4
c_{sj}	4	5	6	6	3	5	6	4	5	4
$g_1(i)$	5	−2	−4	−2	—	—	1	—	—	—
y_i	1	—	—	—	0	0	1	0	0	0
c_{tj}	2	4	2	3	5	6	2	4	3	4
c_{vj}	6	8	8	9	5	7	8	6	8	6
$c^*{}_{tj}$	2	4	2	3	5	6	2	4	3	4
$g_2{}^*(i)$	—	−1	−2	1	—	—	—	—	—	—
y_i	1	0	0	0	0	0	1	0	0	0
c_{rj}	2	4	2	3	3	2	2	4	3	4
c_{sj}	4	5	7	11	5	6	6	7	5	4
$g_1(i)$	—	—	—	1	—	—	—	—	—	—
y_i	1	0	0	1	0	0	1	0	0	0

Let l denote the index realizing the minimum of the right-hand side of (15). If $y_1{}^* = 1$, $c_k{}^* \leq c_{lk} \leq c_{vk}$ as $g_1(l) > 0$ by assumption. If $y_l{}^* = 0$ and $c_k{}^* > c_{vk}$ setting $y_l{}^*$ at 1 and keeping the $y_j{}^*$, $j \neq l$ at their value would reduce $c_k{}^*$ by strictly more than $c_{vk} - c_{lk}$, reduce the sum of the supply costs of the other customers by at least $\sum_{j=1, j\neq k}^{n} \max(c_{sj} - c_{lj}, 0)$ and augment the sum of the fixed costs by f_1; this contradicts the assumption that Y^* is optimal.

Corollary. If

$$g_2{}^*(i) = \sum_j \max(c_{ij}{}^* - c_{ij}, 0) - f_i \quad (16)$$

is < 0, $y_i = 0$ in all optimal solutions of (5)–(9). If $g_2{}^*(i) = 0$, $y_i = 0$ in at least one optimal solution of (5)–(9).

The previous result makes it possible to compute an upper bound $n_i{}^*$ on the number of customers which may be supplied from i in any optimal solution of (5)–(9):

$$n_i{}^* = |\{j \mid c_{ij} \leqslant c_{ij}{}^*\} \mid i \in K_2; \quad (17)$$

if only one optimal solution is sought a strict inequality may be used in (17).

Let us define quantities $c^*{}_{ij}$ for all $i \in K_1 \cup K_2$ and all j by

$$c^*{}_{ij} = \begin{cases} c_{ij} & i \in K_1 \\ c_{ij} + \dfrac{f_i}{n_i{}^*} & i \in K_2, \end{cases}$$

Vol. 197

and quantities c_{rj}^* by

$$c_{rj}^* = \text{Min}\{c_{ij}^* \mid i \in K_1 \cup K_2\}$$

Theorem 2. A lower bound on the values of the solutions of (5)–(9) is given by

$$z^* = \sum_{i \in K_1} f_i + \sum_{j=1}^{n} c_{rj}^*.$$

The proof of theorem 2 is similar to Efroymson and Ray's proof of the validity of the bound of their algorithm. It is therefore omitted here.

Taking (17) into account in the formulae of the additive penalties of Hansen [3] also yields improved bounds. These bounds can be used in direct tests in an obvious way, and combined with penalties, in conditional tests also.

AN EXAMPLE

Let us consider the following example, given by Spielberg [4], the data of which are reproduced in Table 1. Computing the $g_2^*(i)$ by (16) shows 5 variables y_i must be fixed at 0 at the first iteration; then, computing the $g_1(i)$ by (13) shows two variables y_i must be fixed at 1. Computing again the $g_2^*(i)$ makes it possible to fix 2 more variables y_i at 0; finally computing the last $g_1(i)$ shows the last variable y_i must be fixed at 1. The resolution is summarized in Table 2. The example has been solved without any branching occurring. Note that the tests proposed by Efroymson and Ray do not make it possible to fix any variables before branching takes place.

REFERENCES

1. Balinski ML (1964) *On Finding Integer Solutions to Linear Programs*, pp. 225–248. IBM Data Processing Division.
2. Efroymson MA and Ray TL (1966) A branch and bound algorithm for plant location. *Ops. Res.* **14,** 361–368.
3. Hansen P (1972) Two algorithms for the simple plant location problem using additive penalties. Paper presented at the European meeting of the Econometric Society, Budapest.
4. Spielberg K (1969) Plant location with generalized search origin. *Mgmt Sci.* **16,** 165–178.

P Hansen

(*February 1976*)

Faculté de Sciences Economiques
Faculté Universitaire Catholique de Mons
Chaussée de Binche 151
Mons 7000
France

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Sensitivity of the Rate of Return

IN AN EARLIER note [1] Gold's model for managerial control ratios [2, p23] was expressed as

$$r = (p - c)\, ek = \pi cek \qquad (1)$$

where r = rate of return on total investment = profit/total investment;
p = unit price for the output;
c = unit cost for the output;
e = output/capacity = capacity utilisation;
k = capacity/total investment;
π = unit profit relative to cost = $(p - c)/c$.

In order to examine the relative effects of changes in unit price, unit cost and volume on the rate of return, it may be convenient as an alternative to (1) to express r as

$$r = (p - c)v \qquad (2)$$

where $v = e\ k$ = output/total investment. If the unit price changes from p to $p + \delta p$, the unit cost from c to $c + \delta c$ and v to $v + \delta v$ the resultant change in r can be shown to be

$$\delta r = (p - c)\ \delta v + (v + \delta v)(\delta p - \delta c). \quad (3)$$

If the profit margin is denoted by $\alpha = (p - c)/p$ the relative change in r denoted as $r^*(= \delta r/r)$ may be expressed as

$$r^* = v^* + \frac{1 + v^*}{\alpha}\,[p^* - c^*(1 - \alpha)] \quad (4a)$$

which for some purposes may be more conveniently formulated as

$$\frac{1 + r^*}{1 + v^*} = 1 + \frac{p^*}{\alpha} - \frac{1 - \alpha}{\alpha}\, c^*. \quad (4b)$$

Alternatively, these expressions may be written in terms of π (by substituting $\alpha = \pi/(\pi+1)$).

Equation (4) highlights the relative effects of changes in p, c and v. Since $\alpha < 1$ it follows that any given change in p has a greater effect on r than when a change of the same magnitude occurs in c or in v, when $\alpha > 0$; also, if $\alpha < 0{\cdot}5$ then a given change in c affects r more than v does. In most industrial operations $0 < \alpha < 0{\cdot}5$, so that the 'pecking order' (in terms of the expected effect on r) is first p, then c and finally v, but for large profit margins when $\alpha > 0{\cdot}5$ the order is p, v, c, and when α is negative the order becomes c, p, v. Note that for relatively small values of p^*, c^* and v^* equation (4a) may be replaced by the approximation

$$r^* = v^* + \frac{p^*}{\alpha} - \frac{1 - \alpha}{\alpha}\, c^* \qquad (5)$$

and each term in this expression shows the relative effect on r by one parameter when the others are zero (for example, $r^* = v^*$ for $p^* = c^* = 0$, etc.), thereby providing a useful guide as to the expected results for given possible changes in the system and for a given value α.

Equation (4) can also be used to compute the effect on r of several simultaneous changes in the parameters p, c and v. Assume that each may either change by $\pm\Delta\%$ (upwards or downwards) or remain unchanged; the number of combinations of sets of changes in the three parameters is 27, shown in Table 1, where the cases are ranked in terms of the effect on r^* (the first 12 cases showing all the possible combinations for increasing r and the last 12 cases involving a decline). The Table gives the resultant value of r^* for two possible values of Δ (1% and 5%) and for $\alpha = 0{\cdot}1$ and $0{\cdot}2$ respectively. For example, for $\alpha = 0{\cdot}2$ when $p^* = 1\%$, $c^* = 0$ and $v^* = -1\%$ (case 8) the rate of return improves by 4% (all the results are rounded off to the nearest integer).

Table 1 reveals interesting, and in some cases intriguing, results. First, we find that while certain cases are symmetrical (not only when a single parameter is subject to change and where symmetry is to be expected, but sometimes when two parameters change simultaneously, as in cases 2 and 26), in the majority of cases there is no symmetry and increases in parameter values tend to have a greater effect than decreases. Secondly, in some cases the value of α has a significant

TABLE 1. RELATIVE CHANGE (in %) IN THE RATE OF RETURN (r^*) FOR GIVEN CHANGES p^*, c^* AND v^*

Case	p^*	c^*	v^*	$\Delta=$	$\alpha = 0{\cdot}1$ 1%	$\alpha = 0{\cdot}1$ 5%	$\alpha = 0{\cdot}2$ 1%	$\alpha = 0{\cdot}2$ 5%
1	+	−	+		20	105	10	52
2	+	−	0		19	95	9	45
3	+	−	−		18	85	8	38
4	+	0	+		11	58	6	31
5	0	−	+		10	52	5	26
6	+	0	0		10	50	5	25
7	0	−	0		9	45	4	20
8	+	0	−		9	42	4	19
9	0	−	−		8	38	3	14
10	+	+	+		2	10	2	10
11	0	0	+		1	5	1	5
12	+	+	0		1	5	1	5
13	0	0	0		0	0	0	0
14	+	+	−		0	0	0	0
15	−	−	+		0	0	0	0
16	0	0	−		−1	−5	−1	−5
17	−	−	0		−1	−5	−1	−5
18	−	−	−		−2	−10	−2	−10
19	0	+	+		−8	−42	−3	−16
20	0	+	0		−9	−45	−4	−20
21	−	0	+		−9	−48	−4	−21
22	0	+	−		−10	−48	−5	−24
23	−	0	0		−10	−50	−5	−25
24	−	0	−		−11	−52	−6	−29
25	−	+	+		−18	−95	−8	−42
26	−	+	0		−19	−95	−9	−45
27	−	+	−		−20	−95	−10	−48

The change in p^*, c^* or v^* is shown in the second, third and fourth columns as follows:
+ denotes an increase by Δ%;
− denotes a decrease by Δ%;
0 denotes no change.

effect on the results, in others none (as in cases 11, 12, 17 and others). But perhaps the most noteworthy result is that the effect of combined changes in parameter values can be extremely large compared with what appear to be small changes in the parameters themselves. For example, in case 1 for $\alpha = 0{\cdot}1$ when p and v increase by 5% and c decreases by the same amount, the combined effect on the rate of return is to increase it by 105%. In 18 cases of the 27 listed in the Table the changes in r are relatively high, thereby indicating how sensitive r is, particularly when α is small. This result is of some significance both for prediction and for planning purposes, in that it suggests that when forecasts for changes in p, c and v are given in the form of ranges rather than by single values, the resultant change in r is likely to be very variable. This was corroborated in a case study using risk simulation for the analysis of plant performance, where it was found that while the distribution of unit cost was fairly compact for a given set of forecast changes, the corresponding distribution for the return on investment was very wide [2, pp. 141–2].

It should be noted that when it is desirable to decompose the parameter v into several ratios, equation (4) can easily be expanded

Vol. 4
1976

by substituting the appropriate value of v^*. For example, if $v = ek$ as in equation (1), it can be shown (following the method in [1]) that

$$v^* = e^* + k^* + e^*k^*$$

and similarly when v is expressed as the product of three ratios (as in Gold's model in [2]) v^* can be easily found prior to the use of equation (4).

REFERENCES

1. Eilon S (1975) Changes in profitability components. *Omega* **3** (3), 353–4.
2. Eilon S, Gold B and Soesan J (1976) *Applied Productivity Analysis for Industry*. Pergamon Press, Oxford.

Samuel Eilon
(February 1976)

Department of Management Science
Imperial College of Science and Technology
Exhibition Road
London SW7 2BX
UK

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Implications of Goal Programming for Insurance Agency Decision Making

MANAGEMENT decision making in many large insurance agencies and brokerage firms has become a complex task. One of the problems involved in management decisions in these firms is that a variety of goals influence the decisions. These multiple goals are often interdependent and conflicting, thereby further complicating the decision process. This paper suggests the use of goal programming in insurance agency decision making involving multiple, often conflicting goals.

Vol. 4
1976

In order to use goal programming, a brokerage firm or agency first has to determine its goals or objectives. Several goals which may be important to an insurance agency are discussed herein. The goals suggested are neither all encompassing nor constant among agencies. Instead, these goals are selected because they offer a diversification in the type of goals an agency might develop. In any specific case, more goals may be added or perhaps only a subset of the goals will be applicable. One goal of an agency might be to write more business. Although income maximization is a concern to an agency in the long run, growth may be more important to management in the short run. In terms of a goal programming model, this objective can be stated as a percentage increase in premium written in the current year over the previous year.

Another agency goal may be to place the insurance coverage it sells not only with its current insurers but also with new insurers. An agency might feel that utilizing the services of new insurers may be beneficial because of higher commission scales and because of additional underwriting expertise. Therefore, instead of relying mainly on the group of insurers which it currently represents, an agency would want to increase the number of companies with which it places insurance. New insurers would require a minimum amount of insurance premiums. In a goal programming model, this goal would require that some minimum level of premiums be written for each insurer.

An agency might be desirous of balancing its business among insurers, thereby not becoming dependent on any one insurer. In the event an insurer should restrict the business it will accept, the agency would have numerous other sources for placing coverage. This goal would be the antithesis of the previous goal. It would require that there be a maximum amount of premiums that could be placed with each insurer.

Some insurers currently provide services for a firm, such as billing, policy issuance, etc. When these services are afforded, normally there is a reduction in commission scales. To the extent that these services are optional, an agency could be desirous of utilizing any services that offer a cost reduction equal to or greater than the corresponding reduction in commissions.

As mentioned previously, income maximization might not be of primary importance to an agency; however, a firm must have some income to survive. Therefore, income, but not income maximization, could be a goal. Another goal could be maximization of contingency commissions. To include this goal, a brokerage firm or agency would place the most profitable business with the insurer that offered the highest contingent commissions. This goal would differ from the usual profit goal. In the profit goal, the agency would be examining initial commissions, i.e. guarantees. An attempt to maximize contingency commissions is a gamble which may or may not materialize depending on loss experience.

Finally, an agency goal could be to write more commercial and less personal coverage. There are several reasons why this might be a goal. The firm may have expertise in the commercial area which has not been utilized. Commercial business may be easier to handle, or it may be more profitable.

To establish a goal programming model, an

agency or brokerage firm must establish priority levels for the agency goals. The conscious effort devoted to goal prioritization may, in fact, be more beneficial to the agency than the results of any specific goal programming model. After the priority levels have been determined, the above mentioned goals can be represented by numerous constraints at the different priority levels. The authors have had success with problems involving ten insurers and thirty-one classes of insurance in cases with six goals and four priority levels.

Claude C Lilly
John M Gleason
(*December 1975*)
College of Business Administration
Texas Tech University
Lubbock, Texas 79409
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Comment on 'Stochastic Dominance for Ranking Ventures'

THE PURPOSE of this note is to extend the recent work by Tilley and Eilon (referred to henceforth as TE) [3] in stochastic dominance theory. A new rule for the class of decision makers with piecewise linear utility functions is presented. This rule can be used to alleviate the drawback of stochastic dominance criteria described in TE's paper. This rule, however, requires additional information with regard to the decision maker's utility function since the calculation of 'break-even' point is necessary in determining the class of utility functions. Further discussion of this matter will be given at the later part of this note.

TE have pointed out some limitations of stochastic dominance rules as selection criteria (see page 180 in [3]).

The 'inconclusive region' in TE's a–b space characterizes the unfortunate drawback of all three stochastic dominance criteria for any random variable. This follows from two facts concerning the stochastic dominance of F over G:

(1) It is necessary that the difference $G-F$ have its first sign positive.
(2) It is necessary that $E(X) \geq E(Y)$.

Now, if we return to TE's example and consider (a, b) belonging to the 'small inconclusive' triangle, then we note that $E(X) = \frac{1}{2} > a + b/2 = E(Y)$, and $G-F$'s first sign is negative (see Fig. 6 in [3]).

Thus, (a, b) in the inconclusive region implies conditions (1) and (2) cannot hold simultaneously, i.e. F is indifferent to G under first, second and third degree stochastic dominance.

They also note that, for this example, the SSD efficient set is equal to the TSD efficient set. Indeed their result is more general. The corollary to the following theorem shows that the SSD efficient set is equal to the TSD efficient set for any random variable whose cumulative distribution functions intersect only once.

Theorem 1.[1] Let X and Y be the random variables with distributions F and G. If F intersects G at most once, $G-F$'s initial sign is positive, and the mean of X is at least as large as the mean of Y, then X SSD Y.

Proof: Let $T(y) = \int_{-\infty}^{y} [G(t) - F(t)]dt$, then $T(y) \geq 0$ for all $y \in R$. Suppose not, then $T(y) < 0$ for some point $y^* > y_0$, a point of intersection, but then $T(y) < 0$ for all $y \geq y^*$. In particular,

$$\int_{-\infty}^{\infty} [G(t) - F(t)]dt < 0, \text{ a contradiction}$$

Therefore X SSD Y.

Corollary. Under the same conditions as in theorem 1, X SSD Y if and only if X TSD Y.

Proof: The proof is immediate since X TSD Y implies $\int_{-\infty}^{\infty} [G(t) - F(t)]dt \geq 0$, and X SSD Y always implies X TSD Y.

In order to further reduce the size of the efficient set, TE considers the following 'break-even' utility function as a possible candidate for an individual's utility function:

$$U(x) = \begin{cases} c_1(x-v) & x < v \\ c_2(x-v) & x > v \end{cases}$$

where v is a well-defined break-even value for returns.[2]

The following theorem gives a necessary and sufficient condition for the ordering of

[1] This formulation of the equivalence of the SSD and the TSD criteria was suggested by an anonymous referee.

[2] Although TE do not assume anything with regard to the sign of c_1 and c_2, we shall assume that $c_1 \geq c_2 > 0$ and interpret that the 'rate of dissatisfaction' (c_1) is faster than the 'rate of excess utility' (c_2).

random variables through 'break-even' utility functions.

Let $U_v = \{u: u(x) = c_1 (x - v)1_{(-\infty, v)} + c_2(x - v)1_{(v,\infty)}, c_1 \geqslant c_2 > 0\}$

where 1_A is a characteristic function of set A.

Theorem 2. Let X and Y be the random variables with continuous distributions F and G such that $E\,u(X)$ and $E\,u(Y) < \infty$ for any $u \in U_v$. Then $E\,u(X) \geq E\,u(Y)$ for all $u \in U_v$ if and only if the following two conditions hold.

(i) $E(X) \geq E(Y)$

(ii) $\int_{-\infty}^{v} G(X)\mathrm{d}x \geqslant \int_{-\infty}^{v} F(X)\mathrm{d}x$

Proof:

Let $\Delta u = E\,u(X) - E\,u(Y)$.

Then $\Delta u = \int_{-\infty}^{\infty} u(x)\mathrm{d}\,(F - G) =$

$$- \int_{-\infty}^{\infty} (F - G)\,u'(x)\,\mathrm{d}x^{3}$$

$$= \int_{-\infty}^{\infty} (G - F)\,[c_1 1_{(-\infty, v)} + c_2 1_{(v,\infty)}]\mathrm{d}x$$

$$= \int_{-\infty}^{v} (G - F)c_1\mathrm{d}x + \int_{v}^{\infty} (G - F)c_2\mathrm{d}x$$

$$= c_1 \int_{-\infty}^{v} (G - F)\mathrm{d}x + c_2 \left[\int_{-\infty}^{\infty} (G - F)\mathrm{d}x - \int_{-\infty}^{v} (G - F)\mathrm{d}x\right]^{4}$$

$$= (c_1 - c_2) \int_{-\infty}^{v} (G - F)\mathrm{d}x + c_2 \int_{-\infty}^{\infty} [(G - F)\mathrm{d}x]$$

$$= (c_1 - c_2) \int_{-\infty}^{v} (G - F)\mathrm{d}x + c_2\,[E(x) - E(y)].$$

Then, clearly $\Delta u \geq 0$ if (i) and (ii) are satisfied. To show that the conditions are also necessary, we consider the following three cases:

(1) $E(X) < E(Y)$ and $\int_{-\infty}^{v} G(x)\mathrm{d}x \geqslant \int_{-\infty}^{v} F(x)\mathrm{d}x$,

(2) $E(X) \geqslant E(Y)$ and $\int_{-\infty}^{v} G(x)\mathrm{d}x < \int_{-\infty}^{v} F(x)\mathrm{d}x$,

(3) $E(X) < E(Y)$ and $\int_{-\infty}^{v} G(x)\mathrm{d}x < \int_{-\infty}^{v} F(x)\mathrm{d}x$.

[3] We note that the derivative of u cannot exist at v since the utility functions being considered are piecewise linear functions with no value at v. However, the *δ-nbhd* type of arguments can be used.

[4] Here, we have assumed that the second piece of the utility function u has value $c_2(x - v)$ on $(0,v)$. If we assume that the second piece of the utility function has value zero on $(0,v)$, then the condition (i) changes to (i') $\int_{v}^{\infty} G(x)\mathrm{d}x \geqslant \int_{v}^{\infty} F(x)\mathrm{d}x$. This theorem can be generalised to the utility function whose pieces are concave on (v, ∞) and convex on $(-\infty, v)$ (Friedman–Savage type).

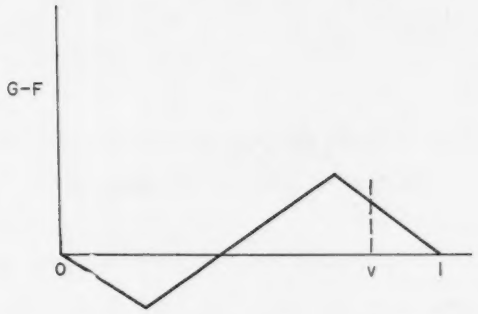


FIG. 1. *Difference of cumulative distributions* F *and* G.

Considering case 1, since $E(X) < E(Y)$, we can choose c_2 large enough so that for any fixed c_1, $\Delta u < 0$, i.e. we can find $u \in U_v$ and $\Delta u < 0$. Thus case 1 and $\Delta u > 0$ cannot hold simultaneously. Cases 2 and 3 can be proven similarly. Hence conditions (i) and (ii) are necessary conditions.

If we denote the above orderings by $\geqslant_v$, then $X(\text{SSD})Y$ implies $X \geqslant_v Y$. The important difference between SSD and $\geqslant_v$ is that $\geqslant_v$ is capable of ordering random variables with $E(X) > E(Y)$ whose initial sign of $G - F$ is negative.

To illustrate the above claim, let (a, b) belong to the 'inconclusive' region in Fig. 6 of [3]. Then the graph of $G-F$'s initial sign is negative as shown in Fig. 1.

Then, as noted above, it is impossible for F to dominate G in first, second and third degree. However, if the 'break-even' point is located sufficiently close to 1, then $F \geqslant_v G$.

At this point, the following comment is in order. The reason why $\geqslant_v$ is able to compare those random variables belonging to SSD efficient set follows from U_v being the proper subset of $U_2 = \{u: u \text{ is an increasing concave utility function}\}$. In fact, $\bigcup_{v \in R} U_v$ can be used to generate the set U_2 [1]. Furthermore, by considering only U_v, where v is sufficiently close to one, we have eliminated from U_2, those decision-makers who insist on $G-F$ be positive everywhere. We also note that $\geqslant_v$ is similar to the chance constrained programming problem where one prefers F to G if $F(T) \leqslant G(T)$ for some target point T [2].

ACKNOWLEDGEMENT

The author wishes to thank Professor Ziemba and anonymous referees for their helpful comments.

REFERENCES

1. MEYER PA (1966) *Probability and Potentials*. Blaisdell, Waltham.
2. PYLE DH and TURNOVSKY SJ (1972) Safety-first and Expected Utility Maximization in Mean-Standard Deviation Portfolio Analysis. Review of Economics and Statistics.
3. TILLEY RPR and EILON S (1975) Stochastic dominance for ranking ventures. *Omega* **3**, 177–184.

D Kira
(*February 1976*)
Management Science Division
Faculty of Commerce and Business Administration
University of British Columbia
Vancouver
BC, Canada

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press. Printed in Great Britain

Optimization Models for Fraction Defective Control Charts—a Markovian Approach

IN A RECENT paper by Ladany and Alperovitch [1] a model for optimal set up policy for defective control charts was proposed. The authors suggested a procedure for computing the values of the sample size, the length of the interval between samples and the control limit that minimizes the expected average cost per unit of time. The purpose of this note is to show that by formulating the model as a Markovian process one can easily extend it to various set-up policies.

Using the notations in [1] let:

P_t = the probability that a shift will occur during an interval of length t;
k = the control limit;
α = the probability of error of type I;
P_a = the probability of error of type II;
n = the sample size;
c_1 = the cost of sampling one item;
c_2 = the cost of not detecting a change in the process for interval of length t (we assume, for simplicity, that if a shift occurs, it is always at the end of the interval);
c_3 = the cost of false indication;
c_4 = the cost of unscheduled set-up;
c_5 = the cost of scheduled set-up.

In [1] the authors assume a model with unscheduled set-ups, i.e. set-up is performed only when the outcome of the sample indicates so. The underlying assumption of the model is that the occurrence of shifts in different intervals are independent events. Thus we can formulate the process as a finite Markov chain with the following definition of states:

state '1'—at the beginning of the interval the system is under control and *immediately* after a set-up.
state '2'—at the beginning of the interval the system is under control and *not* immediately after a set-up.
state '3'—at the beginning of the interval the system is out of control.

The transition probabilities will be then:

$$p_{11} = P_t(1 - P_a) + (1 - P_t)\alpha;$$

$$p_{12} = (1 - P_t)(1 - \alpha); \quad p_{13} = P_t P_a;$$

$$p_{21} = P_t(1 - P_a) + (1 - P_t)\alpha;$$

$$p_{22} = (1 - P_t)(1 - \alpha); \quad p_{23} = P_t P_a;$$

$$p_{31} = 1 - P_a; \quad p_{32} = 0; \quad p_{33} = P_a.$$

From these transition probabilities we can easily compute the steady-state probabilities Π_1, Π_2, Π_3 and find the average expected cost per interval of length t as

$$AC(t) = (\Pi_1 + \Pi_2)[c_1 n + P_t(1 - P_a)c_4 + (1 - P_t)\alpha c_3] + \Pi_3[c_1 n + c_2 + c_4(1 - P_a)]. \quad (1)$$

The average expected cost per unit of time will then be

$$c_n = \frac{AC(t)}{t}. \quad (2)$$

The optimization procedure may then be carried out on (1) and (2). Note that the expected length of a cycle (i.e. the time between two consecutive set-ups) is π_1^{-1} (the term $E(i)$ in [1]). It can be easily verified that the expressions in [1] for the expected cycle length and for the probability of set-up are in error.

The case of scheduled set-ups [1, 2] can be easily constructed using the same approach, the states of the system being

state '1'—the machine is under control and *immediately* after a scheduled set-up.
state 'i'—the machine is under control i periods after a scheduled set-up $i = 2,\ldots,f$
state 'i''—the machine is out of control at i periods after a scheduled set-up.

The transition probabilities may then be stated in a similar manner to the unscheduled set-up cost and the steady-state probabilities can be computed accordingly:

$$\Pi_1 = \frac{1}{f}; \quad \Pi_t + \Pi_{t'} = \frac{1}{f}$$

The average expected cost $AC(t,f)$ for an interval of length t is then found and the average expected cost per unit of time is

$$c_u = \frac{AC(t,f)}{t}$$

which can be optimized with respect to the relevant decision variables.

The previous cases represent two extreme policies. In the first set-ups are never pre-planned while in the second case they occur at intervals of f periods. A more common policy is to combine the two and schedule a set-up when a given amount of time has elapsed since the last set-up (scheduled or non-scheduled). The method proposed in this note can be readily applied and the expression for the average cost can be found as in the previous cases.

REFERENCES

1. LADANY SP and ALPEROVITCH Y (1975) An optimal set-up policy for control charts. *Omega* **3**, 113–118.
2. LADANY SP (1973) Optimal use of control charts for controlling current production. *Mgmt Sci.* 763–772.

Yoram Alperovitch
***Yoram Friedman**
(*March 1976*)

**Faculty of Management*
The Leon Recanati Graduate School of Business
Tel-Aviv University
PO Box 39010
Tel-Aviv
Israel

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 3, 1976. Pergamon Press 1976. Printed in Great Britain

International Register of Theses and Dissertations

The following are available at the Yvette Jacobson Library, Management Centre, University of Bradford, Emm Lane, Bradford, Yorkshire BD9 4JL, UK:

BENSON ML (1972) Interpersonal Communication and Personal Influence in the Menswear Market: A Pilot Survey. MSc Thesis, University of Bradford.

BILLOT HE (1970) An Approach to Manpower Planning in a Large Engineering Organisation. MSc Thesis, University of Bradford.

BIRCHALL DW (1972) Work Attitudes and Job and Organisational Change. MSc Thesis, University of Bradford.

BONNINGTON PJ (1971) The Car Industry and the Common Market. MSc Thesis, University of Bradford.

BRADLEY PG (1971) Management by Objectives: An Appraisal. MSc Thesis, University of Bradford.

BROWN A (1970) Business Budgeting. MSc Thesis, University of Bradford.

CLEGG CW (1972) Warehouse Location Theory: A Critique. MSc Thesis, University of Bradford.

CONNORS MFJ (1972) An Examination of Management by Objectives as a Technique Enhancing the Motivation of Managers. MSc Thesis, University of Bradford.

COOK DE (1970) A Review of Some Methodological Aspects of Diffusion Research. MSc Thesis, University of Bradford.

COOPER S (1971) Production Planning and Control in Footwear Manufacture—A Case Examination and Proposed System. MSc Thesis, University of Bradford.

COWELL DW (1970) Automatic Vending as a Method of Distribution in the United Kingdom MSc Thesis, University of Bradford.

CURRILL DL (1970) The Role of Technological Forecasting in the Integrative Planning of Technological Development: A Study of Six UK Companies. MSc Thesis, University of Bradford.

DAVIES AJ (1972) The Corporate Model and Financial Management. MSc Thesis, University of Bradford.

DAVIS D (1973) The Nature, Classification and Use of some Methods of Mass Production. MSc Thesis, University of Bradford.

DAY P (1972) A Set of Comparisons between various Short Term Seasonal Forecasting Methods. MSc Thesis, University of Bradford.

DOBSON P (1971) An Investigation of Certain Factors Involved in the Marketing of Construction Machinery. MSc Thesis, University of Bradford.

FINNEY MJ (1972) The World Petroleum Industry. MSc Thesis, University of Bradford.

FOULGER BR (1973) Costing and Pricing Decisions, MSc Thesis, University of Bradford.

GIBSON BRC (1973) An Appraisal of Group Technology and its Implications for a Company in the Machine Tool Industry. MSc Thesis, University of Bradford.

GOLD JG (1973) The Giveaway Industry. MSc Thesis, University of Bradford.

Vol. 4
1976

HAMMOND A (1971) Job Shop Forward Loading. MSc Thesis, University of Bradford.

HARKNESS BL (1971) Establishment of an Effective Inventory Control System for a Company in the Machine Tool Industry. MSc Thesis, University of Bradford.

HAWKINS K (1970) The Reform of Collective Bargaining in British Industry. MSc Thesis, University of Bradford.

HAYWARD G (1971) Diffusion of Innovation in the Flour Milling Industry. MSc Thesis University of Bradford.

HERBERT PJA (1972) Underwriting Issues of Corporate Securities. MSc Thesis, University of Bradford.

HILLIER TJ (1970) The Growth and Development of 'Out-of-Town' Shopping Centres in the United Kingdom: A Market Viewpoint. MSc Thesis, University of Bradford.

HILLIER TJ (1972) Decision Making in the Industrial Buying Process. PhD Thesis, University of Bradford.

The following are available at the University of California Library, Berkeley, California 94720, USA:

DE SIMONE LE (1974) A Short-Run Regional Econometric Forecasting Model of Residential Construction Activity: An Application to California Standard Metropolitan Statistical Areas. PhD Dissertation, University of California, Berkeley.

DUDLEY, CL JR (1974) A Theoretical Financial Analysis of the Long Term Subsidy Value of the Regional Development Incentives Program in Canada, and a Related Paper on Corporate Investment Theory. PhD Dissertation, University of California, Berkeley.

HIBSHOOSH A (1974) A Model of Brand Competition in Attribute Space: Toward a Formal Framework for Marketing Analysis of Consumer Behavior. PhD Dissertation, University of California, Berkeley.

MATULICH S (1971) Technical Security Analysis and Portfolio Simulation by Computer. PhD Dissertation, University of California, Berkeley.

MEREDITH JR (1972) A Markovian Decision Model for Evaluating Clinical Programs for the Mentally Retarded. PhD Dissertation, University of California, Berkeley.

NEILSON DP (1974) An Evaluation of Computer-Instruction in Undergraduate Accounting Curricula. PhD Dissertation, University of California, Berkeley.

PRENTICE CA (1971) Anticipatory Financial Accounting—An Empirical Study. PhD Dissertation, University of California, Berkeley.

SMITH JRM (1974) A Model for Data Selection Policies in Information System Design: A Theoretical Approach. PhD Dissertation, University of California, Berkeley.

VERTIGAN MJ (1974) Impact of Alternative Profit Measures on Selected Macro Economic Models. PhD Dissertation, University of California, Berkeley.

WALTER GA (1971) Augmentation of Task Group Performance Through Video Tape Feedback and Social Learning. PhD Dissertation, University of California, Berkeley.

YOUNG IC (1971) The Impact of Deposit Interest Rate Ceilings on the Level of Economic Activity. PhD Dissertation, University of California, Berkeley.

ZIMMERMAN JL (1974) Individual Financial Decision Making in Basic Science: A Normative and Descriptive Study. PhD Dissertation, University of California, Berkeley.

The following are available at the Joseph Regenstein Library, University of Chicago, 1100 East 57th Street, Chicago, IL 60637, USA:

BURMAN GR (1973) The Economics of Discrimination: The Impact of Public Policy. PhD Thesis, University of Chicago.

HENDRICKX NJ (1973) Optimal Production Schedules and Inventories for Some Continuous Dynamic Deterministic Multi-Echelon Models. PhD Thesis, University of Chicago.

HOGARTH RM (1972) Process Tracing in Clinical Judgment: An Analytical Approach. PhD Thesis, University of Chicago.

OUCHI WG (1972) A Novel Approach to Organizational Control. PhD Thesis, University of Chicago.

PAULUS JD (1972) The Estimation of Large Systems of Consumer Demand Equations Using Stochastic Prior Information. PhD Thesis, University of Chicago.

PHELPS CE (1973) Demand for Health Insurance: A Theoretical and Empirical Investigation. PhD Thesis, University of Chicago.

SENGENBERGER W (1972) Industrial Inefficiency and the Incidence of Productivity Bargaining in Great Britain in the 1960's. PhD Thesis, University of Chicago.

SPENCE RJ (1974) Exclusive Dealing—Its Rationale and Effect on Competition: A Case Study. PhD Thesis, University of Chicago.

The following are available at the Library, Cranfield Institute of Technology, Bedford MK43 OAL, UK:

COAD RA (1971) SKINFLAIR—Study of an Entrepreneur. MBA Thesis, Cranfield Institute of Technology.

COLLICUTT B (1972) A Study of Diversification Opportunities for a Small Manufacturing Company. MBA Thesis, Cranfield Institute of Technology.

COMPTON AM (1972) An Examination of the Market Potential for Anodised Aluminium Crossover Coils in Hi-Fi Loudspeaker Units. MBA Thesis, Cranfield Institute of Technology.

COOPER DA (1972) Equipment Leasing Feasibility Study. MBA Thesis, Cranfield Institute of Technology.

CORDINER J (1971) The Strategy for the Market Development of a New Perishable Consumer Non Durable Product. MBA Thesis, Cranfield Institute of Technology.

DAGRES GD (1971) Some Aspects of Applied Reliability Theory. MBA Thesis, Cranfield Institute of Technology.

DAVIES G (1971) An Investigation into the Twenty-One Larger Construction Companies Based in or on the Border of Wales. MBA Thesis, Cranfield Institute of Technology.

DAVIES RR (1972) A Critical Evaluation of the Uses and Benefits of Captive Insurance Operations. MBA Thesis, Cranfield Institute of Technology.

DE ALWIS G (1972) Financing of Industrial Projects in Developing Countries with Special Reference to Sri Lanka (Ceylon). MBA Thesis, Cranfield Institute of Technology.

DE MORAES NA (1970) An Examination of, and Recommendations for, a Small Business. MBA Thesis, Cranfield Institute of Technology.

DIXON JR (1972) A Study of the Significance of Information Relating to a Major Development Decision. MBA Thesis, Cranfield Institute of Technology.

DUFFEY PJ (1972) The Study of a Wage System. Brookhurst Igranic Ltd. MBA Thesis, Cranfield Institute of Technology.

EXHAM RL (1970) Capital Expenditure Control—An Investigation on Behalf of Messrs. Truman, Hanbury, Buxton & Co. Ltd. MBA Thesis, Cranfield Institute of Technology.

FEENEY JA (1972) Marketing Study for a Private Company. MBA Thesis, Cranfield Institute of Technology.

FEHLING CEA (1971) An Investigation into the Market for Financial Advertising and Public Relations with Special Reference to the Place in this Market of Dewe Rogerson. MBA Thesis, Cranfield Institute of Technology.

Vol. 4
1976

FERREIRA JPS (1972) The Application of Some Forecasting Models to a New Product Launch. MBA Thesis, Cranfield Institute of Technology.

FITZ-JOHN P (1970) A Study into the Cost Effectiveness and Responsibilities of the Estimating Staff at a Major British Builders' Merchant. MBA Thesis, Cranfield Institute of Technology.

FITZPATRICK DJ (1970) Hamersley Ore in Europe: A Transportation Study. MBA Thesis, Cranfield Institute of Technology.

FLAHERTY ST J (1970) An Investigation of Customer Service for Texas Instruments Ltd. MBA Thesis, Cranfield Institute of Technology.

FLETCHER DE (1971) A Survey of the Educational Technology Market. MBA Thesis, Cranfield Institute of Technology.

FLETCHER JR (1971) Investigations of New Applications for a Moulding Process. MBA Thesis, Cranfield Institute of Technology.

FOX GS (1971) The Financial Function in an Engineering Organisation. MBA Thesis, Cranfield Institute of Technology.

FOWLES REG (1972) Investigations Carried Out in the Aircraft Works Division of Westland Helicopters Limited, Yeovil, Somerset. MBA Thesis, Cranfield Institute of Technology.

FRENCH CEM (1970) Profit Planning. MBA Thesis, Cranfield Institute of Technology.

FRETHEIM J (1971) The Financial Control System of Det Norske Veritas. MBA Thesis, Cranfield Institute of Technology.

FRY PD (1971) A Survey of Magistrates Courts with Particular Reference to Witnesses Waiting Times. MBA Thesis, Cranfield Institute of Technology.

The following are available at the Lyon Playfair Library, Imperial College of Science and Technology, Exhibition Road, London SW7 2BX, UK:

AUCHELIDES J (1972) LNG Carriers: Construction and Shipping Costs. MSc Thesis, Imperial College, London.

BAKER BM (1972) The Time Constrained Salesman Problem. MSc Thesis, Imperial College, London.

BANTIN DA (1972) Objective Setting for Companies in the Construction Industry. MSc Thesis, Imperial College, London.

BEER CPC (1972) An Investigation into Aspects of Manpower Planning using a Corporative Planning Model. MSc Thesis, Imperial College, London.

BENNETT DG (1972) Recovery and Re-Use of Resources. MSc Thesis, Imperial College, London.

BUNN DW (1972) A Bayesian Formulation for Inference on the Future. MSc Thesis, Imperial College, London.

COLES JM (1972) Forecasting the Sales Consumer Durable Goods. MSc Thesis, Imperial College, London.

COLMAN JG (1971) Planning in a Manufacturing Company. MSc Thesis, Imperial College, London.

COX WP (1971) Mathematical Models for Corporate Planning. MSc Thesis, Imperial College, London.

DALE PM (1971) A Financial Survey of Mergers During the Years 1968/70. MSc Thesis, Imperial College, London.

The following are available at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 30 Memorial Drive, Cambridge, MA 02139, USA:

BEALS AL (1970) Relationships Between Control and Effectiveness in Brazilian Development Banks. SM Thesis, MIT, Cambridge, MA.

BERNHART MH (1970) The Effect of Reference Groups on Responses in Interviews. SM Thesis, MIT, Cambridge, MA.

BEVILLE JJ JR (1970) Merging and Main Stream Control Techniques for an Automated Highway System. SM Thesis, MIT, Cambridge, MA.

CECIL JG (1971) The Evaluation of Computer Systems Assistance to Managerial Decision Making. SM Thesis, MIT, Cambridge, MA.

FRADKIN HE (1971) A Survey of Bank Interest in a Cash Flow Computer Model. SM Thesis, MIT, Cambridge, MA.

GERSTBERGER PG (1971) The Preservation and Transfer of Technology in Research and Development Organizations. PhD Thesis, MIT, Cambridge, MA.

GREEN RS (1971) Heuristic Coupling of Aggregate and Detail Models in Factory Scheduling. PhD Thesis, MIT, Cambridge, MA.

HERZOG EL (1973) Work Relationships in the Delivery of Health Care: An Analysis of the Division of Labor Between Physicians and Nurse Practioners in Outpatient Clinics. PhD Thesis, MIT, Cambridge, MA.

HJELM KE (1971) The Future Role of Consultants in Developing Economics. SM Thesis, MIT, Cambridge, MA.

KLEIN BE (1970) Application of Sharpes's Index Model to Portfolio Theory and Practice. SM Thesis, MIT, Cambridge, MA.

KOTTER JP (1970) The Psychological Contract: Expectations and the Joining-Up Process. SM Thesis, MIT, Cambridge, MA.

LEJEUNE PY (1973) Career Paths: A Study of Sloan School of Management Alumni Sample. PhD Thesis, MIT, Cambridge, MA.

MARCUS RG (1971) The Inventory and Distribution Policies of a Large Industrial Rubber Goods Firm. SM Thesis, MIT, Cambridge, MA.

MCLEAN EH 3RD (1970) A Computer-Based Medical History System: Factors Affecting Its Acceptance and Use by Physicians. PhD Thesis, MIT, Cambridge, MA.

PETERSON SE (1970) Management, Motives and Society—A Study of the Relationship between Sociological Variables and Some Attributes of Personality among the Upper Classes in Iceland. PhD Thesis, MIT, Cambridge, MA.

RANDERS J (1973) Conceptualizing Dynamic Models of Social Systems: Lessons from a Study of Social Change. PhD Thesis, MIT, Cambridge, MA.

ROTHERMEL TW (1970) A Marketing Examination of the Antitrust Definition of the Relevant Market. PhD Thesis, MIT, Cambridge, MA.

SHLENSKY BC (1970) Determinants of Turnover in NAB—JOBS Programs to Employ the Disadvantaged. PhD Thesis, MIT, Cambridge, MA.

SLAVICH DM (1971) The International Financing Decision: A Programming Approach. PhD Thesis, MIT, Cambridge, MA.

SOLNIK BH (1972) European Capital Markets: Towards a Theory of an International Capital Market. PhD Thesis, MIT, Cambridge, MA.

TAYLOR AC (1972) Water Quality Planning in a River Basin. PhD Thesis, MIT, Cambridge, MA.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

EDITORIAL

The Quality of Working Life

MAN HAS always been concerned about work, as a means for earning a living, as an outlet for a creative urge and as a framework for social interaction. These broad purposes encompass many facets of man's needs and ambitions, his desire for material as well as mental rewards, his instinctive competitive spirit and the satisfaction that he derives from recognition of his efforts by his fellow-men. And since the working man spends a high proportion of his waking hours on activities associated with his work, it is not surprising that social scientists and industrial engineers have become increasingly interested in the conditions of the workplace and its environment, or in what is now coined as a general umbrella title *the quality of working life*. One example of this growing interest is the international conference on this theme held in September 1972, followed by two edited volumes of some 30 papers and 15 case studies, published in 1975 [1], where authors from several countries and various backgrounds covered many strands of what is undoubtedly a difficult subject.

Vol. 4
1976

The concern about the quality of working life is largely centred on such questions as:

1. What are the major elements of disaffection experienced by working people and what are their causes?
2. How are man's needs to be ascertained? How do they change with increased material well-being and with personal development? How are they affected by changes in the work environment and in the external environment?
3. To what extent are conditions at work determined by prevailing technology and/or organization structure, and under what circumstances is it possible to change the technology and to specify organizational parameters that will culminate in a desired environment at the workplace?
4. Is there a conflict between economic performance of an enterprise and the quality of working life? What are the reasons for such a conflict, and how can it be resolved?

This is not intended as a comprehensive list of questions related to a discussion on the quality of working life, but it gives a good indication of how broad the subject matter is and why investigators are so tempted to segment it. It is only natural for such segmentation to be largely dictated by well-established disciplines, with the social scientists putting the emphasis on workers' attitudes and the fulfilment of individual aspirations, the industrial engineers on task and output performance, the managerial economists on enterprise performance, the system analysts on information flows and linkages between various parts of the system. Each discipline inevitably starts with its own set of 'givens', and what is 'given' to one is not necessarily accepted and often hotly contested by another. Attempts to gloss over these differences by a so-called multi-disciplinary approach, allowing each discipline to make its own contribution to an integrated and well-structured analysis, often fail for the simple reason that the multi-disciplinary approach tends to become a watered-down compromise, in which the disparities in basic assumptions and 'givens' are rarely reconciled.

What emerges vividly from the questions posed above, is the central issue of measurement. One cannot discuss intelligently ways in which the quality of working life can be improved, or how the possible undesirable effects of certain characteristics of technology and organizational design can be avoided without an explicit measure of the quality of working life and without being able to state whether and in what manner it has changed from one time period to another. Ideally, this problem of measurement should be kept entirely separate from whatever value judgements people might have as to how good or how bad a given working environment is. In other words, it is one matter to be able to assert that a system has changed, or is likely to change, from state A to state B and quite another to register one's approval or disapproval regarding the incurred or proposed change. This point needs to be stressed, since many of the research workers pontificating on the quality of working life start with a declared intention to improve it, and this almost invariably (although it need not be an inevitable outcome) leads them to adopt measurement techniques with an in-built bias, which—while it remains entirely consistent with their own system of beliefs—offends others who start from different premises.

That this is a fundamental issue is evident from the measurement methodologies that have so far been proposed. Walker [1, p.1] warns us that " 'what the worker wants' is not an infallible guide to quality of working life", that "an enterprise is a *coalition* of individuals who may have competing interests", that we need to explore the conditions that influence the worker's attitudes and behaviour, although it is not essential to determine the degree of rationality of such attitudes—all this sounds like a sensible and promising introduction. But six papers later, comprising part of the book devoted to "Defining and measuring the quality of working life", I felt none the wiser.

All the proposed definitions are naturally embedded in their related lists of

criteria or categories. For example, Walton [1, p.91] suggests eight yardsticks: adequate and fair compensation, safe and healthy working conditions, immediate opportunity to try and develop human capacities, opportunity for continued growth and security, social integration in the work organization, constitutionalism in the work organization, work and the total life space, and the social relevance of work life. Apart from the fact that this form of wording requires further elaboration, that the ensuing set of definitions will beg questions of yet further semantic connotations, that the criteria inevitably overlap and in some respects correlate with each other, it is not clear what you do with these yardsticks once their definitions are understood and accepted. If the intention is to score each criterion on a scale, how are the scores to be accounted for: in a grand aggregate measure of quality of working life, by a simple arithmetic addition, by weighting, by an eight-element vector?

That some aggregate measure is contemplated is implied in Walton's discussion of the effect on productivity [1, p.99], illustrated by a graph of productivity (taken as a single measure) plotted against "the opportunity to use and develop human capabilities" (again taken as a single measure, with no explanation as to how it is to be constructed, except the intimation that it is "created by autonomy, multiplicity of skills required, work information, and planning responsibilities"). This graph suggests that maximum productivity is attained at a certain "optimal level of opportunity to use and develop capacities", beyond which productivity proceeds to decline, whereas a graph of "quality of work experience" continues to rise and attains a maximum value at a much higher level of "opportunity to use and develop human capacities", at which point productivity is well below its optimal level. Accepting that this kind of presentation is merely illustrative of a conjecture that there is a conflict between productivity and the quality of working life and that the graphs are not meant to convey any quantitative connotations, there is a clear implication that productivity, the opportunity to use and develop human capabilities, and quality of work experience can each be determined and described by a single measure, and consequently that a relationship between these variables can be found as a basis for some action research. Alas, no empirical evidence is available to support such a measurement methodology, nor even any intimation as to how the necessary evidence is to be gathered.

Other examples fare no better: Seashore [1, p.105] points out that there are three perspectives (that of the worker, that of the employer, and that of society), each with its own set of criteria; he further argues that the testing of a given set of propositions about job satisfaction measures "has distinct limitations", and that less than half the variance of such measures is explained by a relatively small number of environmental conditions (leaving about half the variance attributable to demographic and personality characteristics). Sheppard's four-way paradigm [1, p.121] to take account of physical and economic attributes is supposed to differentiate between sixteen 'types' of quality of work satisfaction

Vol. 4
1976

and he argues that it should be used "to determine the proportion of a nation's (or an enterprise's) work force in each of the categories", and to monitor changes over time, so as to provide the necessary information for the design of "intervention strategies". There is an explicit assertion in this paper that some 'types' are better than others (one is categorized as "ideal", having "good physical work environment, high job security, high wages, and high job satisfaction"), but any ranking exercise in this respect must be regarded as a dubious proposition. Lawler [1, p.129] advocates that in addition to reporting on earnings, companies should report where they stand on a "quality of work life index", assuming that such an index can be agreed upon, and he suggests that "probably behaviorally trained human system auditors would be needed to audit the human system, just as the accountant audits the financial system". The mind boggles.

These and other conjectures regarding possible procedures for measuring the quality of working life are all variations on the same theme: they strive to provide a typology, or an index, based on a number of specified key variables, and far from there being any agreement on which variables to include in the measurement scheme, little thought appears to have been devoted to the sheer mechanics of the measurement process, to the consequent levels of precision and accuracy, to problems of bias and interpretation. The viability and validity of such measurement exercises are usually taken for granted, so that the transformation of theoretical propositions into practice is rarely expected to pose problems of a fundamental nature. Such an attitude is not likely to be shared by experienced field workers, many of whom have come to regard the results of such measurement exercises with understandable scepticism.

Turning from the problem of measurement to motive, the many authors on the subject of quality of working life are united in searching for means to improve it. Herrick and Maccoby [1, p.63] postulate four principles for humanizing work: the principle of security (to relieve the worker from anxiety about safety, income and future employment), the principle of equity (to compensate the worker according to his contribution to the value of the product or service involved), the principle of individuation (to encourage the worker to develop himself, his ability to learn, his craftmanship), and the principle of democracy (to allow the worker to participate in the control of his environment and "'take responsibility for what is produced, how money is invested, and for the social consequences of production'").

That advancement on these four points would be desirable is perhaps a proposition that—in general terms—would find universal acclaim, but such approbation must be coupled with several important reservations. First, the authors assert that the four principles "together describe a system that is constructed to optimize the worker's well-being and, correspondingly, that of society". Apart from the fact that "optimize" is often an over-worked term, whose precise meaning in this context is somewhat in doubt, there is no evidence

Vol.
197

to support the statement that "such a system would develop in the worker a sense of hope, activeness, and productiveness, alleviating symptoms of discontent, mental illness, and despair", far less to suggest that more 'well-being' for the individual is more 'well-being' for society. Many examples have been cited to demonstrate that there is often a conflict of interest between individuals, that a greater level of job satisfaction for one may be attained at the expense of another, that freedom of action for the individual can be incompatible with the interests of the society of which he is a member. It is both futile and irresponsible to ignore such conflicts. Thus it is illusory to suppose that individuation (what a dreadful term!) is possible for all[1], that a better educated work-force must result in the humanization of work, that greater automation for higher productivity will increase the need for better qualified workers and hence improve everyone's level of job satisfaction—all these are explicit or implicit assertions made by the authors.

Secondly, it is far from clear at what rate, and indeed how far up the scale, it is desirable to progress with regard to the four postulated principles. What, for example, is a reasonable level of security? Is it an infinite dimension with new horizons defined as desirable targets at whatever point you are on the scale, or can a fixed datum be determined as a proper balance between the need to relieve individual anxiety and that of economic viability of the enterprise and, if so, how is this datum to be found? Or take the principle of democracy: to what extent can a developing industrial democracy remain compatible with the need to keep in check the process of bureacratization, the need for efficiency in decision-making, and the need for the enterprise to react and adapt rapidly to a changing environment?

Vol. 4
1976

Cited case studies are, alas, not sufficiently revealing to allow judgements to be made about the precise effects of a redesign of the working system, particularly in the long term. Time and time again one finds that descriptions of implemented changes highlight the benefits to some groups of workers but ignore the resultant disbenefits (real or perceived) to others [1, pp.69–70]: In one company all employees were on a monthly payroll and had the same general conditions of work—sounds fine for those who had been working on a weekly basis with shorter holidays, but what about the erosion of differentials and relative fringe benefits for those who carried greater responsibilities? In another company the organization was designed to eliminate the need for supervisors—resulting in clear advantages to working groups, but what about the loss of status and careers for current and would-be supervisors? A third company allowed organized cleaning service teams of nineteen people to be "given a voice in planning, problem solving, and goal setting" (how?), with teams rather than individuals being accountable for work done—but what about those whose personal contribution was thereby less recognized?

[1] I have discussed this topic elsewhere [2].

Again and again authors on this theme emphasize the attributes of work-group decision-making as a means for decentralized control. A typical classification of the degree to which the group can assume responsibility for the managerial process is given by Gulowsen based on experience gained from trials in several Norwegian plants [1, p.75]:

"A. The group members decide their individual production methods.

B. 1. The group decides the internal distribution of tasks.
2. The group decides on questions of recruitment.
3. The group decides questions regarding internal leadership.

C. 1. The group decides what additional tasks to take on.
2. The group decides when it will work.
3. The group decides questions regarding the production method.

D. 1. The group has influence upon its quantitative goals.
2. The group decides questions regarding external leadership.

E. The group has influence upon its qualitative goals."

Apart from the fact that an organization structure based on work groups is thought to be compatible with, if not an integral part of, the goal of industrial democracy, the involvement of the worker in decisions associated with his work place is clearly designed to combat alienation caused by his remoteness from the production plan and compounded by highly repetitive and boring tasks. The group offers him a voice, a means of participation, an opportunity to widen his responsibilities and diversify his activities. No wonder that "group technology" has increased worker satisfaction and moral, has reduced absenteeism, and has consequently lead to improvements in key productivity performance indicators.

But there is a danger that the uninhibited enthusiasm generated by the case studies reported so far could blind us to other consequences of the work group organization. Consider, for example, the possible conflict of interests between groups, regarding which jobs each should have a responsibility for, disparities in production rates and targets, even arguments on quality standards. Since each group may be dependent on others for its inputs and affect others with its outputs, such conflicts can easily occur. And when changes in the working environment take place—changes in product design, retooling, layout, production technology—existing conflicts are likely to intensify.

What is of even greater concern is the position and freedom of the individual within a working group. By making the group the cornerstone of the organization structure, the will of the individual becomes subjugated to the groups. Admittedly, he participates in the decision-making process, and he has an opportunity to influence others; if the group operates as a true and genuine democratic entity, the individual can afford to lose many battles without a

feeling of injustice. But circumstances could arise of jealousy, of power struggles, of strong-arm tactics and group tyranny. In such circumstances the individual could become the powerless victim of a well-intentioned system; his only option then would be to opt-out and thereby lose his job, his security, and all the other desirable attributes associated with the quality of working life. Enthusiasts for new organizational designs would do well to contemplate these issues and to produce an appropriate antidote which it may be necessary to apply when over-reliance on the magical powers of the group fails to produce the anticipated goods.

SAMUEL EILON
Chief Editor

REFERENCES

1. DAVIS LE and CHERNS AB (eds) (1975) *The Quality of Working Life*. The Free Press, New York.
2. EILON S (1975) The Myth of Self-Actualization. *Omega* **3** (6), 623–630.

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

FEEDBACK

A Comment on "A Dynamic Model of Process and Product Innovation"

UTTERBACK and Abernathy are to be commended for their attempt to develop a more comprehensive theory to explain patterns of innovation in industry [3]. They rightly point to the need to explain the central tendencies and systematic variables in the innovation process between firms and different industrial segments. And they illustrate the uses that can be made to this end of previous empirical studies of industrial innovation, such as that of Myers and Marquis [2]. However, we have some doubts about the validity of their model, and of the empirical confirmation supposedly given to it by the data of Myers and Marquis.

The essence of their argument is that "characteristics of the innovation process and of a firm's innovative attempts will vary systematically with the firm's environment and its strategy for competition and growth, and with the state of development of process technology used by a firm and by its competitors" (p. 2).

"Definitive stages of development that are similar in different industries and economic sectors can be identified in the characteristics of the productivity factors of different processes. The pattern of changes from one stage to another in the process are pervasive in character going beyond the physical attributes to the productivity factors themselves" (p. 3).

There are three such stages:

Stage I: uncoordinated process, product performance-maximising;
Stage II: segmented process, sales-maximising;
Stage III: systemic process, cost-minimising.

At each of these stages there are differences in the rates of product and of process innovation, and in the relative importance of market, technology, output rate and cost factors in stimulating such innovation. In Stage I, product innovation is high, is market stimulated, incorporates relatively big amounts of new technology and has relatively high costs. By State III, process innovations are predominant, are stimulated by production-related factors, incorporate relatively little new technology, and are relatively cheap (see Fig. 1, [2]). Utterback and Abernathy rightly point out that their analytical framework is similar to that of economists using product cycle theories to explain patterns of international trade and investment. It is in fact very similar to that employed by Hirsch for his study of the US electronics industry [1].

However, we do not think that the data from the Myers and Marquis study confirm the Utterback/Abernathy model, and for a rather simple reason. Utterback and Abernathy have taken as their sample only 330 of the 567 innovations (i.e. 58%) described by Myers and Marquis.

The reasons for this selection are not entirely clear. On page 10, it is argued that innovations were eliminated from multi-divisional firms whose product lines lacked homogeneity; yet the data in the Appendix definitely do not suggest that these firms were significantly bigger than those included in the sample.

We therefore think that it is instructive to look comprehensively at the data collected by Myers and Marquis on all the 567 innovations. These can be broken down into five product areas. Admittedly, each product area may comprise a number of disparate 'product segments'. But the same may equally be true of Utterback/Abernathy classification of

Vol. 4
1976

TABLE 1. INNOVATION CHARACTERISTICS IN FIVE PRODUCT AREAS ACCORDING TO MYERS AND MARQUIS

	Railroad companies	Railroad suppliers	Housing suppliers	Computer manufacturers	Computer suppliers	Total
Number of innovations	79	125	196	90	77	567
Primary initiating factor (%)						
Market	19	53	52	31	61	45
Production	48	20	29	36	22	30
Technical	32	22	19	22	13	21
Administrative	1	5	1	11	4	4
Total	100	100	100	100	100	100
Nature of innovation (%)						
Product	0	64	71	54	78	58
Component	0	26	15	31	13	18
Process	100	10	14	14	9	25
Total	100	100	100	100	100	100
Degree of adaptation (%)						
Little or none	39	34	19	2	18	22
Considerable	37	38	62	29	43	45
Invention required	24	29	19	69	39	33
Total	100	100	100	100	100	100
Impact on production process (%)						
Original new process	54	14	28	24	12	26
New for firm only	38	14	39	26	20	29
No change	8	72	33	50	69	46
Total	100	100	100	100	100	100
Cost of innovation						
Less than $25,000	39	34	32	16	49	33
$25,000–$100,000	34	36	40	17	20	32
$100,000–$1 million	19	24	22	29	23	23
More than $1 million	8	6	7	40	8	12
Total	100	100	100	100	100	100

innovations according to the primary initiating factor (see pp. 10–11). Table 1 presents this data which compares a number of features of the innovations in the five product areas: the primary initiating factor, the nature of the innovation, the degree of adaptation, the cost of the innovation, and the impact on the production process.

Only one sector out of the five fits into one of the three stages defined by Utterback/Abernathy. Innovations in railroad companies are strongly of a production-stimulated, process-oriented variety, involving relatively cheap adaptations and very little new technology; in other words, railroad companies fit tidily into the author's Stage III. Railroad and housing suppliers have some of the same Stage III-type characteristics, namely, relatively cheap, adaptive innovations. But they also show a very strong market orientation in relation to both the primary initiating factor (market) and the nature (product) of the innovations; in this sense, they are Stage I-type innovations.

Exactly the reverse dichotomy is shown in the computer manufacturing sector. Compared to the average for the five sectors, production and technology are relatively

important as primary initiating factors, and components relatively important in the type of innovations; in this sense, computers are probably a Stage II product. But in terms of the cost and inventiveness of the innovations, computers are the most strongly Stage I product area of them all. Finally, computer supply companies do almost fit completely into Stage I, except that their innovations are relatively inexpensive.

Thus, the Utterback/Abernathy model is not wholly consistent with the different patterns of innovation observed by Myers and Marquis in the five product areas. Even if it was, it can still be argued (as the authors themselves recognise on p. 17) that a cross-sectional analysis of innovations does not show the dynamics of change within any process segment from one stage to the next.

We are doubtful whether patterns of innovation can, as Utterback and Abernathy claim, be explained independently of the specific characteristics of each sector. For example, there are probably very strong reasons specific to the railroad companies, that cause them to concentrate on process innovations: regulatory constraints, physical constraints, and company strategy immediately spring to mind. Similarly, since railroad and housing suppliers make intermediate goods, there are strong pressures on them to improve the quality and the performance of their products, even if the lack of scientific and technical opportunities mean that they tend to concentrate on adaptive innovations. On the other hand computer manufacturers have a much richer scientific and technical base (particularly in componentry) on which to draw in order to make expensive innovations, whilst computer suppliers can make such innovations more cheaply.

Examples from other sectors also suggest that the specific characteristics of particular products, their use and their manufacture have a strong influence on the pattern of innovation, and that this pattern may be relatively constant through time. At one extreme, process innovation has always been important in the manufacture of steel, of bulk plastics and of synthetic fibres. At the other, extreme heavy electrical generating equipment has always been made on a one-off basis and has required continuous and important product innovation.

We therefore conclude that a framework for explaining different patterns of innovation in different process segments must embrace a far larger number of variables than those identified by Utterback and Abernathy; for example, the degree to which product and process technology is science-based; the end use of the product; the opportunities for static and dynamic economies of scale in processes, or of economies of scale in product; the relative size and technical capabilities of suppliers, manufacturers and users; the dynamics of competition; and regulatory and physical constraints. We share the wish of Utterback and Abernathy to be able to move to higher levels of generalisation about innovation processes in industry but fear that, given their complexities and specificities, this will be an arduous task.

Vol. 4
1976

REFERENCES

1. HIRSCH S (1965) The United States electronics industry in international trade. *Natn. Inst. Econ. R.* (34), 92–97.
2. MYERS and MARQUIS (1969) *Successful Industrial Innovations*. NSF, Washington DC.
3. UTTERBACK JM and ABERNATHY WJ (1975) A dynamic model of process and product innovation. *Omega* 3(6), 639–656.

K PAVITT[1]
R ROTHWELL

[1]*Science Policy Research Unit*
The University of Sussex
Mantell Building
Falmer
Brighton, Sussex BN1 9RF
UK

A Policy Lapsation Model

VARDE's paper demonstrates how much can be done towards improved measurement by constructing a model of the situation under review. It is in the concluding remarks that the key to progress lies. More extensive data will allow the lapse rates of different classes of insurance policies to be investigated by duration and by volume. This may give important clues as to the major factors leading to early lapsation. Having established

these factors it will be much easier to alter the method of paying commission so that categories of policy with a high lapse rate receive a lower commission payment. The agent will thus be encouraged to sell categories with a low lapse rate.

In the UK the balance between initial and renewal commission has always provided fertile ground for argument about rewards to agents. We have a self-compensating system where higher commission will always attract more new business but will leave less of the competitive gross premium to build up the necessary reserves. A rapidly expanding life office therefore develops weak reserves.

In India there is presumably less difficulty in gathering the extensive statistics needed to analyse each class of business by lapse rate and duration but in the UK the collating of data from several offices is made difficult by the varying methods of keeping records and a general reluctance to reveal methods or standards to a competitor. There is a primary belief among insurers that any competitor will draw advantage from a comparison with his own performance statistics and that this will exceed any advantage gained by the insurer submitting data for comparison. This desire for secrecy means that it is impracticable to carry out any satisfactory analysis by duration because of shortage of data.

When this work by the Life Insurance Corporation is developed to the next stage it will be of considerable interest to smaller countries like the United Kingdom where individual companies will be able to compare the effects of each cause of lapse with their own intuitive unmeasured ideas.

An attempt has been made in the Transactions of the Faculty of Actuaries Vol. 31 p. 56 to measure the effects of age and duration on lapse rates in spite of shortage of data.

REFERENCE

1. VARDE V (1976) A policy lapsation model. *Omega* **4** (3), 331–338.

P GILES

Scottish Amicable Life Assurance Society
Service and Systems Division
Craigforth PO Box 25
Stirling FK9 4UE
UK

Vol. 197

A Note on "A Linear Programming Model for Determining an Optimal Regional Distribution of Petroleum Products" by C Moore and Z Zoltners[1]

THE VERY interesting paper by Moore and Zoltners [6] exemplifies recent efforts in the United States to apply the tools of Operations Research and Systems Analysis to energy planning. But it cannot be over-emphasized that in the United States, as elsewhere, energy planning is inexorably intertwined with the political process; and, as such, poses some formidable obstacles to the analyst seeking to apply quantitative tools. This is so for three reasons; first, the public, and therefore the political image of economists and econometricians is none too good these days; second, the essentially qualitative nature of political decisions and legislative directives poses some serious barriers to quantitative modelling efforts; and, finally, criteria important to political feasibility and public acceptability—such as ease of administrative implementation and enforcement, or the visibility of equity[2]—are very difficult to formalize within the confines of a quantitative framework.

[1] This note was originally presented as a discussion paper at the International Conference on Regional Science, Energy and Environment, Catholic University Leuven, Belgium, 22–24 May, 1975. These remarks reflect the personal judgements of the writer, and should not be viewed as the position of either the Brookhaven National Laboratory nor the US Energy Research and Development Administration.

[2] In a Federal system of government, inter-regional equity is one of the more important political considerations, with inter-regional trade-offs often spanning several pieces of legislation. See e.g. [4].

In evaluating the utility of the Moore–Zoltners Model, therefore, it is perhaps instructive to recall a few of the problems faced by the United States in energy policy making in the winter of 1973–74, and to examine their ramifications on model-building. One of the major public criticisms of the allocation plans promolgated by the Federal Energy Office in early 1974 concerned the inadequate priority given to the energy sector itself; since the shortage of imported oil placed greater demands on domestic energy producers, it was clear that they needed increased energy for extraction, processing and transportation, a need in conflict with an allocation mechanism based on historic distribution criteria.[3] The latter criterion, of course, has the merits of simplicity, ease of enforcement, and apparent equity. Yet as such deficiencies in the original policy became apparent, and modified on an ad hoc, piecemeal basis, the desirability of incorporating a comprehensive inter-regional input–output model into the overall allocation optimization model becomes clear. Unfortunately that is no easy matter, as the authors have themselves stressed. But even if the effort is judged worthwhile,[4] the additional complexity would tend to distract from a model's practical utility, and, perhaps most important in an area of exceptional political controversy, from its clear interpretation of results. Political decisions (and it should be remembered that energy allocation in times of supply disruption is very much a political matter) result from a delicate and finely drawn balance between competing interests; a simple model, in which policy trade-offs can readily be articulated, may thus be preferable over a complex and perhaps more accurate model. By such a criterion, therefore, the Moore–Zoltners model succeeds well.

Another problem encountered in the American allocation planning experience was the substantial price differential between foreign and domestic supplies,[5] which had an especially disruptive effect on the heating fuel market on the East Coast.[6] Heating costs for neighboring homeowners could vary greatly depending on the source of their supplier—a situation that aroused the sensibilities of the public to equity considerations. In response to the resulting public outcry an FEO proposal in February 1974 called for the seventy-six principal eastern suppliers of home heating oil to have the same proportions of domestic and imported fuel that the average supplier had during the corresponding quarter of 1973. Fortunately, the announcement of the termination of the OPEC embargo against oil shipments on 18 March quickly eased price and availability pressures, making such complex allocation plans moot.

Such are the difficulties encountered when, in the supposed cause of equity, one abandons market forces as a mechanism for resource allocation, and substitutes in its place a maze of governmental regulation, which, however well intentioned, carries the danger of resulting in serious misallocation of resources.[7] The lesson for aspiring model builders, however, is the importance of being able to include equity considerations as formal constraints. The Moore–Zoltners model, for example, could readily be expanded to force equalization of supply in a two tier market; and, as we shall note below, would allow explicit demonstration of the cost of such equity constraints by examination of the corresponding shadow prices.

[3] Allocations to agriculture were also fiercely controversial in view of serious supply disruptions in many commodities markets and concomitant consumer price increases that characterized much of 1973 and 1974. See e.g. [8].

[4] This type of global model, in which an LP representation of the energy sector is iteratively linked to an Input–Output Model, is currently under development at the Brookhaven National Laboratory for the Office of the Assistant Administrator for Planning and Analysis of the US Energy Research and Development Administration; one of its applications is to examine the impact of the introduction of new technologies and inter-fuel substitution on the economy as a whole. For further details, see [3].

[5] Due to price controls, however, there were even sharp differentials in domestic sources (see note 8, *infra*).

[6] For a detailed history of the East Coast experience in this regard (see e.g. [5]).

[7] Indeed, the current investigations into allegations of fraudulent conversion of oil from 'old' sources, whose price was controlled at an artificially low level, into 'new' oil, whose price was significantly higher (in an effort to stimulate exploration and development of new domestic sources) exemplifies the dangers of price regulation.

Vol. 4
1976

It should perhaps be noted that Moore and Zoltners' model has greatest applicability to short planning horizons, say in the order of up to 12 months; longer time horizons would necessarily require a more precise and explicit consideration of inter-fuel substitution effects.[8] This observation, of course, does not detract from the model's usefulness as a policy aid for the type of allocation problem faced by the Federal Energy Office in the winter of 1973–74; if mandatory petroleum product allocation were again to be necessary, the nature of the political process in the United States makes it likely that crisis conditions would already exist, and that allocation programs be initiated at very short notice. Indeed, the most recent experience in the Northeastern United States shows that the extent of fuel substitution was often less than might have been expected; many electric utilities in New York and New England, for example, whose oil fueled generation plants were most affected by the oil supply disruptions of 1973–74, and which technically had the capability to burn coal, did not in fact do so as a result of delays in obtaining the necessary air pollution variances and difficulties in reestablishing coal supply channels.

An important advantage of linear programming as a tool for policy analysis is the interpretation of the dual; since its shadow prices yield direct information on the economic impacts of binding constraints, and may be taken as indicative of the incremental price of a scarce resource that is 'justified' in a competitive market. The determination of such scarcity rents is crucial to policy making for nonrenewable energy resources such as oil, gas, or uranium, and several federal agencies have shown interest in work now underway at Brookhaven for estimating shadow prices in a dynamic LP model of the National Energy Sector.[9]

In conclusion, one might reiterate the judgement that the most successful models are those for which both audience and purpose are clearly, even narrowly, defined. The Moore and Zoltners' model meets both these tests, and consequently, is likely to make a very valuable contribution to the improvement of energy planning procedures.

[8] For an analysis of the importance of inter-fuel substitution and technological change to policy analysis, see e.g. [2].

[9] Information generated by such a model is of interest to federal price policy on natural gas regulation or the allocation of R&D funds to alternative technology development programs. [1] The deficiencies of the marketplace in assigning scarcity rent is well articulated in [7].

REFERENCES

1. Bodin L (1974) A study to assess the application of shadow pricing techniques to national energy resource planning. Report to the National Science Foundation, Brookhaven National Laboratory, Upton, New York.
2. Hoffman KC and Cherniavsky E (1974) Interfuel substitution and technological change. *Energy Modeling Conference, International Institute for Applied Systems Analysis, Vienna, Austria.*
3. Hoffman KC (1974) Coupled energy system—economic models. Conference on Energy Modelling and Forecasting, Berkeley, CA.
4. Howards I and Kaynor E (1971) Institutional patterns in evolving regional programs for water resources management. Massachusetts Water Resources Research Center, Publ. No. 15.
5. Jones HGM, DeJong J, Bronheim H and Nathans R (1975) An analysis of the 1973–74 energy shortage in the New York city region. Energy Policy Analysis Group, Brookhaven National Laboratory, Upton, New York, BNL-20133 to the National Science Foundation, Office of Energy Research and Development Policy.
6. Moore CL and Zoltners AA (1976) A linear programming model for determining an optimal regional distribution of petroleum products. *Omega* **4**(3), 301–311.
7. Nordhaus D (1974) 1974 Report of the President's Council of Economic Advisors; Energy in the Economic Report, *Am. Econ. Rev.* **64**(4).
8. Hearings, Subcommittee on Consumer Economics, Joint Economic Committee (1973) The gasoline and fuel oil shortage. Washington, DC.

Peter Meier

Energy Policy Analysis Group
Brookhaven National Laboratory
Upton
New York
USA

Omega, Vol. 4, No. 4

How Society Works

THE Chief Editor has done a useful service (Vol. 4, p. 1) in drawing attention to and commenting on the publication by OPUS of a booklet entitled "How Society Works".[1] The booklet is part of an exercise to learn how far basic laws and Principles of social science are, or can be, accepted and used. This booklet has no specific authors but is the outcome of contributions by and discussions with very many thinkers over several years. It began with a study made at the University of Pennsylvania which led to the production and wide distribution of a 'working paper' of The Tavistock Institute of Human Relations, London, dated 8 November 1974, but unpublished. In that paper, eleven laws and principles of social science relevant to conflict are identified. To keep the length down and for other specific reasons, a number of important principles were not included. Notable among these were:

(a) how group leaders are chosen (too uncertain) and
(b) the application of game theory (too technical).

However, this working paper was considered to be too heavy for communication to a wide audience and the first action of OPUS on its formation in 1975 was to revise the wording of the context (but not the principles). The text was simplified and the whole made more attractive and digestible.

It is this revised version that is the subject of the Chief Editor's comments. With most of these we agree and I give here only a few return remarks. Professor Eilon queries whether value judgements are really absent. In this booklet this question is largely avoided by wording the principles in a way to apply whatever the objectives. Individuals can be in favour of democracy, dictatorship, growth, no growth, equality, etc. It is claimed that, whatever the view or choice, the principles where relevant will be inescapable. But OPUS, in its name and its objective, is against ignorance—a value judgement if you will.

Professor Eilon also challenges the credibility of the concept of the 'rational man'—human behaviour is not necessarily based on understanding. OPUS's statement that the principles "apply in so far as rationality and logic, rather than emotions, form the basis for individual decisions", is perhaps oversimplified like much else in this highly condensed booklet. Nevertheless, the word 'guarantee' as used in the last paragraph of the Chief Editor's article has really little place in a discussion of the application of social science. All that OPUS claims, both from theory and experience, is that the objectives of individuals or groups are less likely to be achieved if ignorance dominates the decision-maker's thinking.

The effect of education on conflict is admittedly uncertain and the OPUS supposition on this is rightly challenged. For one thing, members of a group may rank other objectives higher than they do that of reducing conflict. In this work the word 'conflict' is usually qualified by an adjective such as 'wasteful'.

The subject of ownership and power is important but it presents considerable difficulty. In the first place, the rules are entirely man-made and vary greatly from one society to another. Secondly, it has proved impossible to find a thorough analysis of the subject. Principles 10 and 11 with their definitions can perhaps act as a beginning for such an analysis.

The Chief Editor's classification of five theoretical approaches to conflict should prove to be useful.

This booklet is intended to provoke thinking and discussion and it is hoped others will follow the lead of the Chief Editor, for which we are grateful.

CHARLES GOODEVE

OPUS
48 *Bryanston Square*
London W1
UK

[1] Available from OPUS (an Organisation for Promoting the Understanding of Society), 48 Bryanston Square, London W.1, UK. *Price including postage, 50p.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Parkinson, Peter and Probability: Implications for Productivity[1]

GERALD NADLER

(*Received April 1976*)

Improving productivity and effectiveness in society and organizations is a need which is not being met satisfactorily in most countries. Goals of society and organizations are defined in circular and spiral terms, one leading to the other, as encouraging the development of individual capabilities, achieving greater effectiveness, attaining a better quality of life, and enhancing human dignity. Parkinson's Law, Peter's Principle and Robinson's Law ("in a bureaucracy, work alters upon advancement") illustrate in satirical form many truths about why organizations and society are not increasing their levels of goal attainment. Clues for overcoming this adversity are the treatment of such laws as probability concepts to demonstrate that future performance of a specific organization cannot be predicted, and the recognition that unstated assumptions—we know *how* to apply accumulated knowledge and the research approach should be followed—need to be converted to much different explicit ones. Evidence demonstrating why this is so and what ought to be done is reviewed. A planning and design approach based on an orientation of purposes–target–results is presented, and some different change principles that result from the approach are described.

Vol. 4
1976

THIS paper argues that the approach that society and organizations use for developing specific solutions to fit specific needs in specific circumstances, significantly affects the quality and quantity of results (e.g. effectiveness and productivity). Simon [18] calls such specific solutions 'the artificial', or artifacts, organizational structures, and other socio–technical systems. The paper focuses on the approach, the process, the modes of thinking, the planning and design methodology, which one follows in the process of developing a specific 'artificial'. For example, state the process you would follow with these real needs:

The 33 member Hospital Council of a large metropolitan area commissions a project to improve nurse utilization as a means of relieving the shortage of nurses. As project director, what process would you follow?

A junior high school wants to develop a program for individualizing education for each student. A committee of parents, teachers, students and administrators

[1] Based on the Fifth Annual Paul Naor Memorial Lecture, 1975, given at the Faculty of Industrial and Management Engineering, Technion, Haifa, Israel.

is selected to develop the program. As chairman of the committee, what approach would you suggest the committee should follow?

The filler cap on the tube leading to the gasoline reservoir in an army tank must admit air when the tank is on land and exclude water when fording a river. It is, as yet, quite expensive and difficult to manufacture. As project engineer, what methodology would you follow in redesigning the cap to eliminate or minimize these problems?

A large city has received funds to plan the revitalization of 200 blocks in its decaying inner city. As planner, what approach to the problem would you recommend?

Your responses are quite likely to be similar to those I have heard many people give: what are other junior high schools doing? Make a study of what nurses are doing now. Get all of the facts about the filler cap materials, manufacturing processes, costs, etc. Make an inventory of skills, houses, factories, etc. in the inner city. Select the nursing departments most severely affected and analyze them to find out what is wrong. Review all the literature on individualization. Call companies offering filler caps for sale to determine what they have to offer. Set up a survey of residents within the 200 blocks to determine what they think the problems are. Call in a computer manufacturer to give you the details on an information system that can manage all of the data needed for individualization.

These responses are based on the conventional approach. That is, they reflect a thought process based on what could be called a poor unstated assumption contained in all levels of the educational establishment: if people learn accumulated facts, then they will automatically know how to convert them into practical results. An unstated corollary follows from this, primarily because of the age of science in which we live: if the research approach is so successful in solving problems of developing generalizations, laws and theories about any phenomenon, then it is the approach to follow in solving planning and design problems.

Two very serious deficiencies of these assumptions are now quite clear. The solutions being generated are qualitatively and quantitatively poor or are unimplemented, and the human efforts needed to generate plans and designs are excessive and poorly utilized. That is, society is not producing innovative, effective, and implemented plans and designs in real organizations, nor is society obtaining the benefits it should from the huge professional and societal resources committed to developing the artificial, as any type of social audit would disclose. The deficiencies are so pronounced, as each of us knows, that two international associations and many people are now seeking viable assumptions and approaches. We challenge, in other words, why society assumes that the thinking processes suitable for developing generalizations are correct for planning and design. It is worth noting that these assumptions are a major contributor to the 'truths' found in 'Laws' of the Parkinson and Peter type, even though they are

Vol.
197

satirically presented. Sufficient results are now available to suggest different assumptions and a different approach.

To adequately cover these ideas, I will discuss

1. the nature of the goals we seek in our organizations, both profit and non-profit;

2. the perversion of these goals, illuminated by the well-known and new satirical laws;

3. the significant clues of probability and unstated assumptions as means to building a different foundation;

4. a new approach that produces positive results that defy the 'Laws'; and

5. some revised change principles for achieving societal and organizational goals.

Vol. 4
1976

Society, as a reflection of the needs and aspirations of individuals, seeks to achieve several goals, realizing that a maximum amount of each is an impossibility. These broad statements probably reflect each of our scales of value regarding such goals:

1. Encourage development of individual capabilities—each person has abilities and intelligence, should seek and can obtain positions to utilize his capabilities, should have opportunities for advancement.

2. Achieve greater effectiveness in all its forms—greater productivity, increased efficiency, improved operational performance (profits, lower sales prices, lower costs per service, etc.), greater quality of products and services, improved utilization of natural resources, etc.

3. Attain a better quality of life, again in all its forms—peace among nations, minimization of conflicts at all levels, higher standards of living, improved health care, better transportation, reduced fatigue in all work activities, satisfactory levels of leisure time, better physical environment, and

4. Enhance human dignity—individual freedom, equality of humans, liberties of thoughts and values, all without harm to or 'using' others.

What is not so readily acknowledged is that these four goals are circular, even spiral in nature. That is, it is not just a matter of optimizing among or calculating trade-offs between the goals, but also that one leads to another. Encouraging the development of individual capabilities is a worthwhile goal in itself, but what is its goal, what is its result? Achieving greater effectiveness. But what is the goal of greater effectiveness, what is its result? Attainment of a better quality of life is the only reasonable response. But, again, what is its

goal, what is its result? Enhancement of human dignity. And, asking the questions for the last time, what is its goal, what is its result? Encouraging the development of individual capabilities is the answer which completes the idea of circularity of goals.

An upward effect along the spiral is what society seeks, and is a suitable definition of what society means by 'progress'. Society, and each organization within the society (and each division, etc.) must decide the rate of progress in attaining various levels of each goal. Toffler [19], for example, suggests the rate of change in the technological component of the goals is too high and increasing, thus causing severe dislocations in the remaining, human-oriented goals. I believe the recent emphasis on the quality of work life, job enrichment, transcendental meditation, transactional analysis, and related movements are manifestations of what society is seeking to do to counteract these dislocations.

Organizations of all types are the structures in which humans exist and are the focal points through which societal goals need to be achieved. A company must achieve all four goals; a government agency, a hospital, a family, a school and so forth must achieve all four goals if the society is to achieve all four goals. To the extent that organizations do not achieve the goals, the society cannot achieve them. And enough folklore and statistical evidence exist to show that our organizations are often neither effective nor human oriented.

This illustrates my second point, the perversion of these goals. Robinson [15] feels "that people may welcome or even actually crave . . . laws and models as impersonal explanations for both good, but especially bad, personal and institutional happenings. Apparently, it allows one to shift responsibility for something from self to the inexorable operation of a natural law. . . . The variety of social laws is surprising . . . and they can be classified as those that are flippant and purely for fun, those that are serious, and those that are satirical, that is, humourous commentaries on institutional and personal shortcomings". The first group is illustrated by Gumperson's Law ("The outcome of a given desired probability will be inverse to the degree of desirability"), Murphy's Law ("If anything can go wrong, it will"), and O'Reilly's Law ("If there is a way to do something incorrectly, that is the way it will be done").

The second classification, serious laws, is illustrated by the Mathew Effect in science ("Them that has, gets"), the Carey law, ("The laws of physical science are equally those of social science. . ."), the organizational entropy laws (one of which is "any human organization will become increasingly disorganized unless work is expended to reduce the tendency toward disorganization"), and Stans' law ("Good budgeting is the uniform distribution of dissatisfaction").

The third category, satirical laws, best captures the organizational perversion of societal goals. Parkinson's Law [12] states that "work expands so as to fill the time available for its completion", and is offered as an explanation of why bureaucrats reproduce themselves at a prodigious rate regardless of the name of the organization—union, government, manufacturing company, education,

health care, and so on. Of course, such fertility has very little to do with the purposes of an organization or department. The Peter Principle [13] claims that "every employee tends to rise to the level of his incompetence". The incompetent often holds back the development of good people and takes credit for what others do.

Robinson's Law [15] states, "in a bureaucracy, work alters upon advancement". After promotion, a person needs to show that his promotion was deserved and another is merited, so he begins to change things. As certain work needs to be performed anyway, and additional work is "required to overcome the inertia of the previous operational routine", the total work becomes greater than that needed for providing the department's services. The change is supposed to produce sufficient productivity improvement to overcome the last two types of work increases, but "experience tells us that this is largely theoretical". What is worse, such changes almost always introduce greater complexity. Perhaps this is what initially motivates the individual because pay scales usually reward increased size and complexity rather than the achievement of purposes!

Vol. 4
1976

Robinson's Law explains Parkinson's Law because, just assuming constant productivity levels, let alone typical lowered one, growth must occur just to produce required outcomes. In other words, current efforts to change require additional people, typically doing less meaningful work, which is just the opposite of our goals. A larger hierarchy creates more opportunities for people to be promoted to their level of incompetence. Put another way, Robinson's Law shows "that if a bureaucracy is limited to a constant size, then its total productivity must steadily decrease." Roman Gaius Petronius expressed the idea in A.D. 66 in this form: "We tend to meet any new situation by reorganizing, and a wonderful method it can be for creating the illusion of progress while producing confusion, inefficiency and demoralization." [4]

Although we can challenge some of the logic of these laws, there is more than enough evidence to claim they protray many facts of real world organizations. Publications [14] assessing the bicentennial status of the United States portray how the country suffers from too much government and, surprisingly coming from the authors who are in the classification, from too many leftish intellectuals. What a perversion of our goals!

But my third point, the clues of probability and unstated assumptions, provides the relief Robinson seeks: "Clearly, some kind of feedback mechanism to constrain the operation of Robinson's Law must be devised so as to bring bureaucracy to a steady state at some level. Without it we are doomed". Again, the satirical nature of his pronouncements should not blind us to the crucial turning point at which he notes that society finds itself. Bureaucracy increasingly *is* stifling, productivity improvements increasingly *are* nonexistent, the citizen and customer increasingly *are* damned for interfering with activities of functionaries, and civil liberties increasingly *are* adversely affected. Should we let these laws continue to operate, or seriously seek to change the 'truths' they contain?

Only the latter course of action is conceivable. It may fail, but without the effort, we do ourselves the injustice of not using our knowledge and intelligence.

Each law must be viewed in probability terms, as a statement of a particular value or range of values in a probability distribution. Because we also know organizations that have demonstrated how better productivity and effectiveness can be attained, we can assign the organizations described by the laws to the low end of the probability distribution whose scale of measurement is "attitude of openness toward improving levels of productivity and effectiveness". Assuming that roughly 15 per cent of the organizations are in the worst or lowest group described by these laws and 15 per cent in the best or highest, 70 per cent of the organizations could be said to have good immediate potential of improving productivity and effectiveness.

Even so, such probability distributions aggregate data describing past conditions and activities of many organizations. Two basic difficulties thus arise: (1) There is no way of identifying where on the measurement scale a specific organization is located, and (2) predicting the performance of an organization based on a particular location in the distribution assumes its future proclivities for change are the same as its past. This is a static rather than dynamic view of organizational performance and it does not accord with known fluctuations in performance over time.

Either one of these alone is sufficient to prove that no law can provide certainty in predicting the performance of any single organization or of any single situation (excepting perhaps death and taxes). Something does not go wrong in every instance where something could (Murphy). Those who do not have anything now sometimes do get something (Mathew effect). Work in some specific organizations does not espand and significant productivity improvements do occur (Parkinson). Not knowing exactly where in the distribution the organization was in the past combines with assumed future sameness to cause people to use the law as a predictor for an organization.

This is patent nonsense. The negative attitude engendered by such laws is perhaps the greatest danger they present. They force people to have an unfavorable attitude towards work thus contributing directly to the lack of satisfactions people find in work. They are viewed as prescriptive, describing what is going to occur anyway, thus detering even the 'good' organizations. They become self-fulfilling prophecies. The negative 'why try?' becomes a means of maintaining the status quo and encourages the willingness to just drift along, or even to revolt as the only way to combat the laws.

What the probability distribution idea does is establish that each organization is unique, is at a different point than any other organization concerning its past and concerning its future ability to change towards the goals and is always capable of more than it is currently attaining. The approach to changing an organization, even the ones that are considered good, must start with where it is located and move it as far as it can go in the direction of the goals. Small

amounts of improvement once every quarter are most often far more effective in reaching the goals than attempting to promote, often unsuccessfully, one big improvement once every 2 years. The smaller changes build a syndrome of success that then augers well for the bigger changes that may occur more often.

The unstated assumptions, the other clue, are not supported by a wide variety of evidence. Here are a few illustrations.

People poorly aggregate and utilize additional information when it is presented in the process of making a decision [3]. Yet the research approach when used for designing the artificial causes people to embark on an information overload basis by collecting huge amounts of data which they could not possibly absorb.

Collecting large amounts of information generates conservative decisions and often leads to the expression, "the more I learn about it, the righter it becomes" [5]. How likely are teachers to adopt individualization if vast amounts of data are collected about their current methods and results?

Psychologists show that people become defensive about their work when shown (challenged is probably a better word) information from conventional data gathering forays that 'demonstrates' something is wrong with it. How many nurses and hospital administrators would placidly agree with the data from the usual nurse utilization studies that show an 18 per cent 'inefficiency' here, a 12 per cent overlap of duties there, and general ineffectiveness everywhere?

Vol. 4
1976

A comparative study [8] of how outstanding planning and design professionals (selected by their peers) approach their planning and design projects found that they do not follow the conventional approach they had been taught. They rejected it because of their intuitive feeling it hampered their creativity. Excellent engineers are likely to approach the redesign of the filler cap differently than the average engineer. Why continue to teach all engineering students such poor approaches?

Another study [20] compared several types of group processes, including the usual interacting mode, and found nominal groups more effective. Would work expand to fill the time available if effective group processes were used with citizens, government officials and related experts meeting about the revitalization planning for the inner city of Milwaukee? Such studies would indicate 'no'.

Is the correct problem being solved [6]? The probability of such an 'error of the third kind' is increased with the conventional approach which forces designers to focus on what exists by analyzing and subdividing it. Is nurse utilization the right problem? I will note later how it is not.

Many other studies could be detailed for you. An experiment involving health planners showed a purpose-based approach produced far better technical results than conventional approaches [11]. Another involving environmental designers showed that a purpose-based approach took less time than con-

ventional ones to produce similar results [7]. A survey of nearly 50 manufacturing companies showed that function and purpose approaches which involved people at all levels and was unconventional produced twice as much yearly cost control savings per staff member. Many comparisons of two different approaches in real organizations when both are inadvertently used simultaneously but independently on the same project show the conventional approaches are less effective.

As such widely divergent evidence makes clear, a root *cause* of the phenomena partially explained by the laws is the conventional approach for designing the artificial. Analysis and subdivision generates 'insight' into details of what exists. What exists must then be done, needed or not, more accurately and/or completely, thus expanding work to fill the time available. As people become experts in the details, more of which continue to exist, they are promoted out of their expertise and often become incompetent. And, because they became experts by analyzing and subdividing, they now repeat the conventional approach seeking to recapture their glory, thus altering work upon advancement.

A synthesis of such knowledge can be structured as a different approach, my fourth point. It is based on a different, but explicit, assumption and corollary. The basic assumption should now be obvious: 'If people want to continually attain higher levels of societal goals, then they must learn a planning and design approach for converting their accumulated knowledge into practical results.' The corollary is then dramatically different: 'If a planning and design result (the artificial) is desired, then a purposes–target–results (PTR) approach will significantly increase the quality and quantity of implemented solutions.' Notice that the research approach remains intact for developing generalizations and theories. Another treatise could be written about the indiscriminate advocacy of additional information in our world, but the question here is not whether we need more information. The question is how best to utilize the vast amounts of information already available and to become available.

I believe that humans have more than one purposeful activity to achieve and that a different approach, or inquiring system as Churchman [2] calls it, is needed for each. I believe there are five others [8, 9], but getting immersed in them here is too much of a digression. My point should be clear if the distinction between research results and planning and design results is established.

The PTR approach involves five integrated ideas:

(1) *The strategy to follow in seeking a solution.* The direction of thinking is the key in its phases, followed in each of the cases presented earlier:

1. Develop a hierarchy of purposes for the stated problem from which is selected the needed purpose the solution should achieve.

2. Generate creative and ideal solutions that achieve the selected and bigger purposes in the hierarchy.

3. Devise the target ideal solution for regularity conditions to guide the detailing of a recommendation. Regularity conditions are those that most frequently occur or are considered most important.

4. Detail the workable solution or system which incorporates the irregularity and exception conditions, while staying as close as possible to the target.

5. Implement the workable solution. Purposes and targets play a critical role in guiding all the minor decisions needed during any implementation.

The hierarchy of purposes is critical for minimizing errors of the third kind. The strategy is flexible and interative, incorporates rather than avoids the unpredictability of 'problems', and focuses on achieving the greatest amount of 'progress' for a necessary purpose. PTR results are pluralistic and multi-channeled. Why throw out what is good for 80 per cent, the regularity, just because the other 20 per cent does not fit? This leads to the multiple channels. But why also force the other 20 per cent, the irregularity, into the 80 per cent channel? This leads to a pluralistic view of solutions.

(2) *A prescriptive framework in which the content of any solution can be described.* A framework of eight elements to be described in five dimensions is utilized in PTR, and can also be considered a universal definition of the word 'system' [9]. It helps determine all aspects of a solution, even in complex systems, insuring that important considerations will not be omitted and possible consequences will not be ignored.

(3) *The effective involvement of people in following the strategy.* A designer or planner is a catalyst and coordinator, stimulating often diverse people with different value sets who will use the results to be involved in developing effective solutions. Their involvement at consultative or supportive levels with an open-ended purpose and target solution orientation generates willingness to accept and implement the results. Other people, such as experts and top decision makers, get involved as necessary. Experts, for example, are important in the second phase of generating solutions.

(4) *Role of organized knowledge, models and techniques.* The PTR approach involves people within the group who know and can utilize the latest technology and levels of knowledge. But the techniques are applied only as the strategy and group decide they are appropriate for a particular project. Technique advocates should not seek problem areas where the techniques can be applied, nor should equipment specialists seek to install a computer, for example, just because it is available. In addition, knowledge and techniques are most often used normatively, to describe what the solution ought to be, rather than to analyze, say with a flowchart, an existing system.

(5) *A program of continuous change and improvement within the organization,* 'Change is stability' characterizes today's society, and each organization,

community, region, etc. should establish a program dedicated to *search* for change. Yet people often act as if a solution in itself is sufficient. This attitude leads to a deadly cyclic pattern, of patching up old solutions as needed until a crisis 'suddenly' emerges requiring large amounts of energy and time and yet another patching. The PTR strategy emphasis on a target solution for a needed purpose should break this cycle, especially if the organization has a planned program of periodic redesign and betterment of all previous solutions, regardless of how well they are currently performing. Critical in the program are efforts *before* any work habits are ingrained to plan and design systems to operate effectively. Establishing a PTR program involves many important policy decisions—how to provide financial return to people, where to locate the program in the organization, how it should operate, how to audit the program performance, and so on.

PTR results are most encouraging. The nurse utilization project, for example, became one of designing a patient care system as a result of expanding purposes into a hierarchy. The implemented system allowed nurses to handle 48% more needed patient care functions with high levels of work satisfaction and quality of care. The system represents 'true' utilization because it is organized around an ideal or target concept of what nurses should be doing. The system was so unique that it won in 1972 a second place Lambert Award out of 1073 entries for innovations in health care delivery. It was so effective that the new 500 bed hospital being constructed by the hospital where the pilot unit was established uses the new system. The filler cap problem, as a second example, was part of a national competition with around 50 entries. The winner followed the PTR approach in his work, and the new design was implemented in the field. What is more, all the comparative studies I noted previously show the PTR approach produced better and implemented results. Much more development is needed, but the direction is extremely promising.

My fifth and final point, some revised change principles, brings together goals, Parkinson, Peter, probability, explicit assumptions, and the PTR approach into concepts which can improve productivity and overall effectiveness.

Change principles can be organized in several ways, but I prefer the timeline scheme. The timeline is immutable, thus presenting the key basis for understanding change. What is done tomorrow and the next day significantly affects the chances of reaching the glorious vision or snapshot of what ought to be done or the target solution. Thus the categories of change principles I will use: getting an organization started, the process of developing the specific changes, and implementation of solutions within a context of continual search for change of all the solutions.

Here are some change principles for getting an organization started:

1. A change structure and attitude must be developed for each organization to fit its uniqueness and overall purpose, goals, and objectives. A currently 'successful' program is likely to be followed by a deterioration in performance

if such a development effort is not started. In addition, if each president of three companies tells me independently each wants a productivity improvement program, I would not tell each that I have just the program (organizational structure and location, type of staff, number of people, types of problems, etc.) each needs just because I had worked with other companies before and will have two other companies setting up programs at the same time. The top management should meet to take part in designing with the PTR strategy the specific program for the specific company. Emphasizing the uniqueness of each organization helps to convince key decision makers *not* previously interested in improvement programs to develop an attitude favorable to and the disposition to take action for seeking real changes in the organization.

2. A solution should be designed for a problem, starting with the first phase of the strategy, in each organization even though every factor in one organization is considered identical to another where an outstanding solution was formulated for an 'identical' problem. The people in each are always different! Assuming that a solution which works in one situation will work in a similar one illustrates the severe dangers of the proof-by-analogy advocates. One hospital with which I worked designed an excellent hospital medical records system, one which received the accolades of peers. Six months later I was working with a second hospital, and one system they wanted designed was medical records. The whole PTR approach was followed with the completely different people and a different purpose emerged and a different system was designed and implemented. Where components in B were the same as in A, then the people in B were delighted to use details developed elsewhere for *their* system. After all, why reinvent the wheel? Just to be sure that a wheel is needed as a component. Schumacher [17] illustrates this principle with the intermediate technology idea of not 'selling' advanced solutions to developing countries. Notice the difficulties which occur with the typical proof-by-analogy reasoning: If the USA uses powered conveyors, fork-lift trucks, and numerically controlled machine tools to achieve high productivity in manufacturing plants, then India needs them to improve its productivity.

3. Words should be adapted to the usage within the organization. Jargon of our disciplines should not be inflicted on organizations. Clear distinctions in ideas, so necessary to finding good solutions, can still be established with the different words as used in the organization.

4. The perceived needs of the organization and its people must form the starting point of any program or project. This adds to the first two principles the important idea of starting with the problem or need the people perceive, even if the potential change based on that need appears to be less than what the change agent thinks could be achieved with other projects. People will change from their current views to a more open view [1] with the PTR strategy. The hierarchy of purposes is particularly adept at expanding the horizons of people to find what needs really exist. There is no reason, for example, to tell a manu-

Vol. 4
1976

facturing manager that his request to reorganize the Industrial Engineering department is not the best problem for him to work on. Start there with purpose expansion so the manager is placed in a situation where he could easily change his perception of his needs.

Notice that in all the principles for getting started, an analysis of what exists is not required. Where the organization is is in the heads of the people to be involved and they do not need it written down. The PTR approach is thus needed to attain these benefits.

Now for the second group of principles, concerning the process of developing the specific changes:

1. The PTR approach should be followed in each specific project needing a specific 'artificial' outcome. It offers a much higher probability of developing a significantly better result than possible with other presently available approaches.

2. Because no organization can have a perfect program of change, the program should generate movement towards (however small) the perfect program. These should be applauded since they are the harbinger of later expansion of change activities.

3. A syndrome of success in holistic or systems thinking should be generated with any size project. The people become more comfortable with larger and more complex problems as a result of their exposure to the PTR systems thinking on previous projects of smaller scope.

4. The change agent must exhibit the same behavior the agent wants a group or individual to adopt. You do not get others to practice participation if you only *lecture* about participation. Using the PTR approach with top management, a change principle already noted, is an important way of getting top management to use it with others.

5. Milestones must be established for all projects and program activities. This avoids dawdling and 'information overload', and forces priorities to be established for the projects.

6. Changes must always be justified. Information and measurements must be available at the start to show the effects of changes. If not, planning such information systems could well become a top order priority.

The third group of principles concerns implementation of solutions within a context of continual search for change of all solutions.

1. Planning and change attention must be committed to all three conditions in which a solution or system can exist—nonexistence, satisfactory existence, or unsatisfactory existence. Original, planned betterment, and correction designs should be continually obtained in the organization [10].

2. Implementation of a solution can itself be considered an 'artificial' specific solution that needs to be designed with the PTR approach. If previous change principles were appropriately used, this design activity is simple. Understanding of people (starting from where they are) is necessary, as the change principles indicate, to implementation [16].

3. Involving people in a continual search for change provides many of the satisfactions from work itself that people seek. The PTR approach particularly helps to do this because of its expansion and ideal concepts.

4. Achieving all four societal and organizational goals is always dynamically uneven. Redressing the balance requires a continual search for change and improvement that should be embodied in a program.

The focus of these change principles provides another antidote for the Parkinson–Peter–Robinson Laws. An expanded number of people to serve as change agents is really not needed. When the unstated assumptions are operative, additional people will be needed, as 'predicted' by Robinson's Law. People now in the organization can, with the PTR approach, become involved in changes and become in many cases the change agent. A seventh grade teacher with no new assistance achieves marvelous results with over 250 students a year each utilizing the PTR approach to design individual study plans in language arts.

In conclusion, I have tried to show that

goals of society and of all organizations, formal and informal, are compatible and constitute the basic measures of progress;

organizational progress leaves a great deal to be desired, as the satirical Laws demonstrate;

probability concepts and the unstated assumptions are clues to changing this behavior;

explicit assumptions and the PTR approach build on these clues; and

effective change principles that generate progress can be stated to take advantage of these developments.

The prospects are encouraging for a new mode of thinking and a reversal of the tendencies that ignore our goals. Liberals in the United States, for example, now recognize that more governmental activity is not always the best way for attaining all of the societal goals. What may initially pass for government intervention to develop human values both does not achieve this goal and does severely reduce the effectiveness of overall government. Managements of many firms are also starting to recognize that they are responsible through their policies and practices for taking the actions which lead to change, and the 'obstinacy' of the unions is not the stumbling block. Unions can adopt these approach ideas too—they are quite large bureaucracies. Responsible union leadership is now saying the same thing.

Each of you is confronted daily with problems that need the design of the artificial. How will you respond? It is the methodological response you and I

Vol. 4
1976

give that will have significant impact on the quality and quantity of goals achieved by our organizations and society, not the Laws that mock human capabilities in satirical cloaks of truths from the past. Still, we can all extract from the past, as I hope this paper shows, those needed concepts to forge ahead. Dedication to questioning the approach you follow would, I believe, be the best memorial each of us everyday could provide for Professor Naor.

REFERENCES

1. ARGYRIS C and SCHON DA (1974) *Theory in Practice: Increasing Professional Effectiveness.* Jossey–Bass, San Francisco.
2. CHURCHMAN CW (1971) *The Design of Inquiring Systems.* Basic Books, New York.
3. EDWARDS W, PHILIPS LD, HAYES LW and GOODMAN BC (1968) Probabilistic information processing systems: design and evaluation. *IEEE Trans. Sci. Cybernet.* **4**, 248.
4. *The Jerusalem Post Magazine.* 21 November 1975, p. 8.
5. MITROFF II (1974) *The Subjective Side of Science.* Elsevier, Amsterdam.
6. MITROFF II and FEATHERINGHAM TR (1974) On systemic problem solving and the error of the third kind. *Behav. Sci.* **19**, (6), 383–393.
7. MURTHA DM (1973) A comparison of problem-solving approaches used by environmental designers. PhD Dissertation, Dept. of Industrial Engineering, University of Wisconsin, Madison, USA.
8. NADLER G. (1967) An investigation of design methodology. *Mgmt Sci.* **13**(10), B642.
9. NADLER G (1975) Highlights of the PTR approach to planning and design. Working Paper, Department of Industrial Engineering, University of Wisconsin, Madison, USA.
10. NADLER G (1970) *Work Design: A Systems Concept*, Ch 18. Irwin Inc., Homewood, IL.
11. NUTT PC (1974) Design methods research: an experimental comparison of the effectiveness of planning procedures. PhD Dissertation, Dept. of Industrial Engineering, University of Wisconsin, Madison, USA.
12. PARKINSON CN *Parkinson's Law and Other Studies in Administration.* Houghton Mifflin, New York.
13. PETER LJ and HULL R (1969) *The Peter Principle.* Morrow, New York.
14. See, for example, *The Public Interest* (1975) No. 41.
15. ROBINSON AH (1975) A variety of laws, to wit: Parkinson's Murphy's *et al.* Dept. of Geography, University of Wisconsin, Madison USA.
16. SCHULTZ RL and SLEVIN DP (Eds.) (1975) *Implementing Operations Research/Management Science.* Elsevier, New York.
17. SCHUMACHER EF (1973) *Small is Beautiful.* Harper & Row, New York.
18. SIMON HA (1969) *The Sciences of the Artificial.* MIT, Cambridge, MA.
19. TOFFLER A (1970) *Future Shock.* Random House, New York.
20. VAN DE VEN A (1972) An applied experimental test of the Nominal, Delphi, and interacting decision making process. PhD Dissertation, School of Business, University of Wisconsin, Madison, USA.

ADDRESS FOR CORRESPONDENCE: *Professor G Nadler, Department of Industrial Engineering, University of Wisconsin–Madison, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Pathways to Management Numeracy

DAVID G LETHBRIDGE[1]

ROGER MILES

(*Received October 1975; in revised form March 1976*)

Development of understanding of mathematical concepts and facility with quantitative techniques as aids to decision-making are important aims for post-experience management courses. Unfortunately, many British managers have a low level of numeracy and find this area difficult. From practical experience with such courses, it is argued that some of the special difficulties and the widely varying needs of middle-aged managers might be met by introducing a self-scheduling learning scheme. The features of this approach to teaching are discussed against a background of the widespread interest in individualized learning in other areas of education. A proposed system for teaching quantitative techniques is described to illustrate this approach and raise a number of design issues.

Vol. 4
1976

INTRODUCTION

MANAGERS are essentially decision-makers; to make effective decisions they must be able to analyse and interpret numerical data and make forecasts against an uncertain future. Unfortunately many British managers have an uncomfortably low level of numeracy compared with their counterparts in other countries and thus are ill-equipped to master the mathematical skills associated with the quantitative aspects of management science.

This situation arises because many present-day British managers lack any background of higher education, and the basic mathematical training which they received at school was generally confined to arithmetic, geometry and algebra. Very rarely have they had even an elementary introduction to the statistical concepts, such as probability and sampling, which are now recognised as vital to sound management practice. Recent changes in school mathematics courses should improve matters in the future but meanwhile the current generation of managers needs considerable help with quantitative methods.

This problem has been met whilst teaching quantitative techniques on post-

[1] Now at the Department of Management Studies, The University of Hong Kong.

experience educational programmes for managers, who typically are aged between 30 and 50 years. One programme, the Senior Managers' Development Programme, has been running for 10 years and has also revealed some of the learning difficulties of the older student. During that time the teaching approach has undergone some significant changes and other improvement possibilities have emerged. For example, the recent arrival of cheap electronic calculators has allowed further discussion of the concepts to replace lengthy number manipulation sessions.

Lethbridge [6] discusses the value of the calculator's contribution and identifies features relevant to the choice of a calculator for classroom use with post-experience management students. He also emphasises that an important change has been to relate teaching more closely to the skills which managers should possess and avoid treating numeracy as an abstract subject. Thus the ability to analyse and interpret numerical data and make sound decisions takes precedence over computational skills. Thirdly the desirability of focusing on the student and his learning processes, rather than on the teacher as a presenter, has been identified as an important principle for future change.

This paper discusses the characteristics of such a system, a student self-scheduling learning scheme, and the benefits which might accrue from its use. There is a growing interest in student-centred and individualized learning schemes in many areas of education, which can be traced back to innovations such as the Personalised System of Instruction, or Keller Plan, which was developed in the USA in about 1962 [5]. Reviewing the considerable advances made by such systems in universities in this country, Boud *et al.* [1] concluded that "it is undeniably one of the most popular innovations that has been proposed in higher education in the last 10 years". Such an approach seems particularly suited to the post-experience management course because of the wide range of needs and abilities exhibited by students on such programmes.

Vol.
197

MATURE STUDENTS AND THEIR RANGE OF ABILITY

The abilities of older students entering a post-experience programme vary enormously. These differences are essentially of four kinds:

1. Large variations in their existing knowledge of specific academic subjects and established disciplines. This is especially true in relation to quantitative methods courses, where classes on entry show a wide spectrum of mathematical ability and skill, ranging from 'left-school-at-14', up to practising operations research managers and statisticians.

2. Considerable variations in experience. Industrial managers are not homogeneous either in terms of their personal backgrounds or in terms of their current jobs, although there are some general skills which might be said to constitute the profession of management. The different functions of management, e.g. finance, marketing, personnel, industrial relations and production—tend to have different technical requirements and provide different types of experience.

3. Variations in the pace at which managers are able to learn. In general, the older the manager, the less able he is to grasp any new concept and therefore the slower the most appropriate learning pace.

4. Differences in the most appropriate learning methodology. Some students appear to learn more readily from visual displays, some would seem to derive most benefit from discussions, while others appear to be able to learn most easily by listening.

Vol. 4
1976

There are also some special problems faced by senior people when it comes to classroom-based learning. Many managers are formally seen as partly playing a teacher's role within their own organisation by virtue of their senior position within the company hierarchy; consequently they may find it especially difficult to revert to the role of the student. In some cases the fear of exposure in front of their classroom peers inhibits the learning process. This seems to occur most often in the quantitative methods area, where the concepts are precise and where the man can be clearly seen to be wrong. Senior managers can sometimes spend more time protecting their own position than they do attempting to learn the subject in question.

A different point, made by managers themselves, is that there are considerable social pressures against classroom questions, because one is acutely conscious of the need not to waste the time of course colleagues. Hence the slower student is loathe to interrupt a session which he is failing to understand, and his learning problems quickly become cumulative. This is especially serious with mathematical topics that have a hierarchical structure so that advanced concepts depend on earlier ones.

Although mature students present the kinds of problems listed above, they do also have their own strengths. The mature student tends to be purposeful and goal-oriented, rather than a seeker for mere diversion. He does not wish for classroom anarchy, nor does he wish to exercise excessive pressures on the teaching staff; he looks for personal achievement and is prepared to work hard. This capacity for self-discipline implies that the mature student will be able to shoulder the enlarged responsibility that he must take if he is asked to schedule his own development programme.

MANAGEMENT DEVELOPMENT PROGRAMMES AND MEASUREMENT OF ACHIEVEMENT

It is usually felt inappropriate to give senior managers public examinations, and one consequence of this is that they often lack any real test of attainment after sitting through an educational programme. A second reason for the absence of quantifiable achievement indices is often a lack of consensus concerning the educational objectives for a specific subject or course. Many areas of management education are difficult to specify in terms of hoped-for outputs; they differ according to the viewpoint taken by the company who sponsors the student, or by the student himself, or the teachers and the educational institution concerned.

However, the quantitative methods area does appear to show a more uniform consensus concerning possible quantifiable objectives, for example in terms of the necessary level of statistical knowledge, than is the case in an area such as marketing. This is important because self-scheduled learning schemes tend to rely on a pattern of achievement tests to regulate progress and to indicate areas of weakness to the student. Whilst these tests need not be public events, they may provoke anxiety in some managers. Certainly students' acceptance of the testing is more likely to be obtained once they are convinced of the value of the whole learning system.

Features of a self-scheduling system

A wide variety of such systems have been produced, often without close adherence to the principles suggested by Keller and the other early pioneers. Yet most, if not all, developers would probably claim the fundamental quality of their system to be giving a student greater independence and responsibility in his learning. Compared with more traditional instructional methods, students are given four types of freedom which Boud *et al.* [1] identify as follows:

1. *Freedom of pace*—a student can work at his own pace and learn at the times and places that he chooses.

2. *Freedom of method*—a student can choose the learning methods that he finds most suitable.

3. *Freedom of content*—a student can choose what he learns within a particular part of the course, or even not to use the course provided at all.

4. *Freedom of assessment*—a student can choose whether or not he is to be assessed, and if so, when and how it should be done.

The extent to which each of these freedoms is available in any one course will depend on many things related to the particular circumstances of that

Vol.
197

course and the characteristics of the institution in which it is offered. For example it may be difficult to give much scope for variations in assessment in courses aimed at externally awarded qualifications. Also when real choices are available, students may be cautious towards the novel ones. Diamond *et al.* [2], describe how students on individualized courses at Syracuse University can choose from a wide range of assessment methods but, interestingly, they find that 50% opt for the traditional written examination paper.

Development of a successful individualized instruction system requires, as described by Gagne and Briggs [3], attention to three points:

1. The development of appropriate learning material. Since these systems involve a shift from the teacher's voice to print and other media as the primary means of instruction, effort must be available for development of the resources.

2. A method for assigning tasks and monitoring progress. The need for an assessment system was mentioned earlier and is important. However, fundamental to the whole scheme must be some basis for agreeing on the learning contract which the student is to undertake. This can be written in terms of explicit learning objectives produced from an analysis of the subject-matter and the job for which the student is preparing.

3. Training for the teacher. A student self-scheduled scheme involves a considerable role change for the teacher, something for which he needs preparation. Now, instead of being the centre of the situation, providing information by lecturing and controlling the classroom, he is an adviser who offers a library of resources and is available for consultation on request. Some teachers will not enjoy the change but others could come to appreciate its merits.

Individualized instruction systems typically give the student more frequent and comprehensive information on his progress than he can gain from conventional classroom tuition. It was this feature which stimulated much of the early interest in the method. For example Keller [5], following Skinner's behaviouristic learning theory, emphasised the importance of such feedback as a provider of reinforcement in the learning situation. Whilst it is clearly important, the feedback provision need not be the dominating feature of a system and supporters of other schools of psychological thought may be attracted to it for other reasons. Shotter [7] argues for treating man "as an active agent able to make things happen, able to construct events", not as a passive reactor to external circumstances. This would lead one to welcome a self-scheduling learning system in which the student has a high degree of responsibility for setting his learning goals and selecting how he is to approach them. Certainly that rationale seems attractive when one remembers the type of atmosphere which is generally desirable on a senior managers' course.

Vol. 4
1976

By allowing the student to select his own educational objectives and required level of achievement, by allowing him to select his own learning media and his own pace of education, by providing an ever-present opportunity for individual feedback on his progress, we would hope to overcome many of the learning difficulties experienced by mature managers. Quantitative methods seems a good topic to start with for the reasons discussed earlier.

The proposed system

The aim is to help the student find a personally scheduled path which will help him to achieve stated educational goals in the shortest possible time and under the best possible conditions. The general benefits should be increased depth of comprehension and speed of learning, enhanced student motivation, a mutual acceptance by both the teacher and the taught of common educational objectives, and more accurate knowledge of the individual student's rate of progress.

The proposed method confines itself to a consideration of only the quantitative methods components of management development programmes. The objectives for which modules could be developed might include:

1. Basic numeracy.

2. Facility with statistics. This could include specific subjects such as: handling data and presenting information, measures of central tendency, measures of dispersion, the normal distribution, sampling theory and sampling practice, hypothesis testing etc.

3. Competence in aspects of operations research. This would cover topics such as: linear programming, decision analysis, network analysis, input–output analysis, dynamic programming etc.

The depth of knowledge to be gained would need to be part of the statement of learning objectives. Decisions as to exactly what objectives are to be offered would be made by teaching staff after questioning students, their host organisations, other teachers and national organisations and government bodies concerned with management education. Once the basic set of objectives for quantitative methods has been decided it will be necessary to devise appropriate measures of attainment.

Assembling the library of alternative learning resources will be a considerable task; where possible it will obviously be desirable to use existing published material which has proven educational value. All the resources will need to be matched to the objectives and tests of attainment before viable learning paths can be established. Thus a particular programmed text might be classified as 'relevant to objective *C*, takes about 2 hr, should be used before test 4 by students capable of sustained individual concentration'. Naturally experience

Vol.
197

will reveal gaps in the bank of resources and indicate where modifications are needed.

Control of a student self-scheduling learning system can involve the teacher in a lot of record keeping. He will need a file on each student which tracks the actual path taken and records progress. There is also a need for a resource control file and a system for booking scarce items. Bottlenecks and delays will occur unless inputs and facilities are scheduled carefully.

Students will no doubt be slightly apprehensive of this new type of course organisation and will need considerable help before they can use it confidently and effectively. In particular, some will find the responsibility of choosing learning objectives and content very daunting if they have been used only to deciding when they should study and for how long. Thus a student guide book must be provided.

Let us suppose that a student, in consultation with his company and the teaching staff, wishes to attain objective *D*-2. His current level of attainment on that objective would be established by a self-administered test or some other form of assessment and he would then be guided to an appropriate level of study. There he would be faced with a range of learning opportunities related to his needs, from which he would select his learning method. The options available might include an audio–visual programme, a written exercise, a practical task, group session, tutorial, lecture or film. Having completed the learning task he could then choose to be re-assessed in one of a variety of ways. The choices would include a written test, an oral examination and submitting a piece of work such as a report or an audio–visual presentation.

Vol. 4
1976

After completing an objective at a particular level the student would then move to another objective or perhaps tackle the same one at a more advanced level. In either case, he would be faced with further choices of learning methods. It is important to emphasise that these would not all involve individual activities; he might well join others for a lecture, tutorial, film or group session. The value of group methods for certain aspects of management development is well known and the self-scheduling system would have to provide for them where appropriate.

A student would move through the options available until he had completed his chosen pattern of study in quantitative methods. Within the whole course some parts would probably be more popular than others, so that a number of students might travel along together for a time, but overall it is quite possible that each student on the course will have created his own unique pathway.

Depending on the numbers of students and the complexity of the course, it may be worthwhile to use a computer-based course management system. A number of computer managed learning systems are now in use, or under development, in education and training in the United Kingdom [4] and are proving useful in situations which offer individualized learning.

A difficulty which must be overcome in designing the course and integrating

it into the teaching institution is that students will spread out and complete their work at different times. Thus if only a part of the whole programme is self-scheduled, some students who finish early will have to wait for the others before they can progress to a conventionally taught phase. It may be necessary to provide some enrichment activities to occupy the faster ones profitably. However, if the whole course were self-scheduled then the early finishers could return to their companies early. Since the mature student is usually paid by his company whilst on course and an expensive replacement must often be provided during his absence, such a speeding up of the learning process could be attractive to companies.

DISCUSSION

Three general observations seem necessary to complete this paper which, since it is essentially speculative, cannot have a firm conclusion.

Firstly, whatever form of independent learning scheme is being developed, it must provide guidance and help for students who lack good study habits. The Open University has had to put considerable effort into helping its students with 'learning to learn'. Much can and should be done to help students organise their learning most effectively and efficiently.

Secondly, despite the great interest in these types of course, there are a great many unknowns requiring research. It is by no means certain which media will be most appropriate in every case and the range of learning styles which students may display is not closely defined. Similarly, some aspects of assessment methodology are controversial. Thus the designer of a student self-scheduling course can learn much from the expertise of others, but in progressing his particular innovation he must be prepared to travel through some relatively uncharted territory.

Thirdly, and lastly, we would suggest that teaching methods which stimulate and enhance the student's ability to organise his own learning have an important general value. Self-reliance will be an important personal attribute for individuals who require continuing education throughout life if they are to play their full part in our rapidly changing society. It would be unwise to assume that formal courses in teaching institutions can meet everyone's needs for new skills and knowledge throughout their working lives. To root this point in our specific context we should recognise that management science is continually evolving new methods and refining good practice. Therefore a manager who wishes to keep abreast of latest developments in quantitative techniques must continually face new learning tasks. There are obvious advantages for him if he can take responsibility for organising his own learning effectively and use scarce and expensive educational resources efficiently.

Vol.
197

REFERENCES

1. BOUD DJ, BRIDGE WA and WILLOUGHBY L (1975) P.S.I. now—a review of progress and problems, *Br. J. Educ. Technol.* 6(2), 15–34.
2. DIAMOND RM, EICHMANN PE, KELLY EF, HOLLOWAY RE, VICKERY TR and PASCARELLA ET (1975) *Instructional Development for Individualized Learning in Higher Education.* Educational Technology Publs, Englewood Cliffs, NJ.
3. GAGNE RM and BRIGGS LJ (1974) *Principles of Instructional Design.* Holt, Rinehart and Winston, New York.
4. HOOPER R (1975) *Two Years On—The National Development Programme in Computer Assisted Learning.* Council for Educational Technology, London.
5. KELLER FS (1968) Goodbye teacher. . . . *Journal of Applied Behaviour Analysis* **1**, 79–89.
6. LETHBRIDGE DG (1975) The Oxford way with managers. *The Times Higher Education Supplement.* 28.11.75, p. 25.
7. SHOTTER J (1975) *Images of Man in Psychological Research.* Methuen, London.

ADDRESS FOR CORRESPONDENCE: *R Miles, National Development Programme in Computer Assisted Learning, 37/41 Mortimer Street, London W1N 7RJ, UK.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Adaptive Trading Rules for Commodity Futures[1]

TERRENCE F MARTELL[2]

(*Received October 1975; in revised form February 1976*)

This paper extends and refines an adaptive filtering model utilizing entropy-based decision rules. The initial adaptive model chose filter sizes by considering both current profitability and previous forecasting success, where forecasting success is measured by a statistic derived from information theory. Alternative specifications of the historical information vector are employed to examine the sensitivity of various components of the model. Although the original adaptive model was a substantial improvement over existing models, it lacked stability; that is, there was more stability in the underlying information of the process than there was in the net profits produced by the adaptive model. Hence this further study.

The information-based adaptive model we developed was run on a large data base composed of 15 years of soybean and wheat data. Our results allow us to reject the Martingale hypothesis of commodity price movement.

Vol. 4
1976

1. INTRODUCTION

THIS paper attempts to reexamine and modify the buy-and-hold (BAH) performance criterion for filter tests of the efficient market hypothesis of commodity price movement and to extend and refine an adaptive filtering model using entropy-based decision rules.

In meeting this objective, we extend the concept of an information-based adaptive filtering rule presented in a recent paper [1]. Although the original model was a substantial improvement over the existing literature, it lacked stability; that is, there was more stability in the underlying information of the process than there was in the net profits produced by the adaptive model. Hence, this further study.

[1] An earlier version of this paper was presented at the 1974 National Meetings of ORSA/TIMS.

[2] The author wishes to acknowledge the financial support provided by the Research Grants Committee of the University of Alabama.

2. THE MARTINGALE HYPOTHESIS AND THE COMMODITY FUTURES MARKET

The now familiar Martingale model can be stated as:

$$\tilde{X}_{(j,\,T+1)} = P_{(j,\,T+1)} - E(P_{(j,\,T+1)} \mid \Phi_T) \tag{1}$$

where $P_{(j,\,T+1)}$ = price of security j at time $T+1$

$$\Phi_T = P_{(j,T)},\ T = 1, 2, \ldots$$

and if prices follow a Martingale process then

$$E(\tilde{X}_{(j,\,T+1)} \mid \Phi_T) = 0 \tag{2}$$

where the tilde over the X designates it as a random variable and Φ_T represents the information set available at time T. If we allow the processes to have normal accumulation, we can then define a submartingale model as:

$$E(\tilde{X}_{(j,\,T+1)} \mid \Phi) > 0. \tag{3}$$

In our case, Φ contains information on past prices of two commodity future contracts, September wheat and September soybeans.[3] Thus, given Φ, we define the 'weak' form of the Martingale hypothesis.[4] For commodity markets the Martingale model (equation 1) is appropriate because it implies that a filter model, however defined, must produce profits greater than zero if it is to find meaningful dependence. It is this standard that is employed in this paper.[5]

3. ENTROPY TRADING RULES, PROFIT AND STABILITY

Having established a criterion for evaluating the efficient market hypothesis as it applies to the commodity futures market, we can now employ that criterion in evaluating the performance of an adaptive filtering model.

Since Alexander's [2] initial contribution, there have been a variety of filter-type security-trading models employed in tests of the Martingale hypothesis. However, with the exception of a recent paper [1], all of these models of investor behavior have one very significant limitation: once a filter size is chosen, that filter size remains invariant across time. This limitation precludes the adjustment of the filter size to new information that might become available over time.

[3] The data base covers the 15-year period from 1956 to 1970.

[4] For a discussion of the 'weak' and 'strong' forms of the Martingale see [3].

[5] This new interpretation of the efficient market hypothesis means that the existing work [1, 4, 5] will have to be reinterpreted.

To counter this limitation, we developed an adaptive filter model,[6] and, following Murphy's [6] contribution, constructed an adaptive decision process of the form

$$S_{T+1} = S(S_T, D_T, E_T)\ T = 0, 1, \ldots, N \tag{4}$$

where

S_{T+1} is the *a posteriori* structural state vector;
S_T is the *a priori* structural state vector at stage T;
E_T is the structural environmental vector;
D_T is decision vector which equals

$$D_T = D(S_{T-\tau}, I_{T-\tau}, V_{T-1})\ T = 0, 1, \ldots, N \tag{5}$$

where

τ is a discrete lag;
V_T is the environmental vector of the decision maker;
I_T is the historical information which equals

$$I_T = I(I_{T-1}, S_{T-1})\ T = 0, \ldots, N. \tag{6}$$

Vol. 4
1976

The heart of the adaptive filtering model is D_T, the decision vector. Here, stored past information and the effects of the environment on past decisions combine to produce a current decision. In this case, the decision vector chooses a filter size for each period for the adaptive model.[7]

To establish the feasibility of an adaptive model, we must relate its components, as defined in equations (4–6), to the commodity market. Thus, the objective of the adaptive model is to achieve the maximum profitability of the process over a fixed time period equal to the life of the commodity contract. Since the structural system is the marketplace for commodity futures, the decision-maker environmental vector contains a measure of the effects of the environment on the decision-maker's action. Thus, the vector contains some measure of the net profits generated by the adaptive filtering model, and the historical information vector contains informational measures of the forecasting success of the various filter sizes employed.

Since the information measure we employ in this paper is the relative average information gain of an increase and decrease forecast, we must develop this measure further and discuss some basic concepts borrowed from information theory.[8]

We define the information content, $h(x)$, of a definite message as the logarithm

[6] A general outline of the adaptive model can be found in Martell and Phillippatos [5]; a more complete discussion can be found in Martell [4].

[7] The specifics of the simple filter model can be found in [7, p. 47]. We are using a variant of $X\%$ filter rule originated by Alexander [2].

[8] For a complete discussion of information theory and some of its applications in economics, see Theil [8].

of the reciprocal of the probability. That is, if x is the probability of an event occurring, then

$$h(x) = \log \frac{1}{x}.$$

We define the information gain of a message to be

$$h(x_0) - h(x_1) = \log \frac{x_1}{x_0} \tag{7}$$

where

$h(x_0) =$ the information content of some event occurring;
$h(x_1) =$ the information content of some new piece of information about the event.

What is of concern to us here is the number of times the filter size of our $X\%$ filter model correctly forecasts commodity price movements. Each filter size forecasts or predicts an increase, decrease, or no change in price, which, after some arbitrary length of time passes may be verified by comparison with the actual price change. The realized price change, thus, may be positive, negative, or zero. A three-by-three prediction–realization table can be developed, then which summarizes the results of our filter's predictive ability.

TABLE 1. PREDICTION-REALIZATION TABLE (PROBABILITIES)

	Realization			
Prediction	Increase	No change	Decrease	Marginal
Increase	P_{11}	P_{12}	P_{13}	P_1
No change	P_{21}	P_{22}	P_{23}	P_2
Decrease	P_{31}	P_{32}	P_{33}	P_3
Marginal	$P_{0\cdot1}$	$P_{0\cdot2}$	$P_{0\cdot3}$	1

where

$$P_i = \sum_{j=1}^{3} P_{ij}; P_j = \sum_{i=1}^{3} P_{ij}.$$

In order to measure the information gain that we received from the predictions we have made, we must define *ex ante* and *ex post* probabilities. Thus, in the case of a forecast increase, we define

P_1 = the percentage of the time an increase was predicted;
P_{11} = the percentage of the time an increase was predicted and did in fact occur;

$P_{0\cdot1}$ = the *ex ante* probability of an increase;

$\frac{P_{11}}{P_1}$ = the *ex post* probability of an increase.

The information gain of an increase forecast, therefore, is

$$h(P_{0\cdot1}) - h\left(\frac{P_{11}}{P_1}\right) = \log \frac{P_{11}}{(P_{0\cdot1})(P_1)}. \tag{8}$$

Proceeding in a similar manner, we can compute the information gain of a decrease and no change forecast.

For comparative purposes, the information gain measure should be standardized by dividing equation (8) by its maximum value $h(P_{0\cdot1})$ which produces

$$\frac{h(P_{0\cdot1}) - h\left(\frac{P_{11}}{P_1}\right)}{h(P_{0\cdot1})} = \text{relative information gain.} \tag{9}$$

The maximum value of equation (9) is 1, which occurs when P_{11} equals P_1. Before discussing how a measure of the relative information gain is used, we must first examine the method for computing such a measure.

In order to compute the relative information gain for a particular filter size, it is first necessary to develop a prediction–realization table for a particular filter size. To use the relative gain concept, we must first develop a prediction–realization matrix with enough observations to be meaningful. This can be accomplished by building a sufficiently large data base before starting our adaptive model. Since we had 15 years of daily observations on both September wheat and beans, we used the first 2 years to develop our initial prediction–realization matrix. The remaining 13 years, in addition to serving as the data base for formal tests of our adaptive filter model, also are used to update continually our original matrix. This is necessary if we are to monitor continually the predictive ability of the filter sizes of our filter model.

Thus far, we have developed the concept of information gain and defined how we compute and interpret this measure. The result of these computations is a relative information gain for each filter size and forecast. What we are really interested in, however, is the relative information gain of an increase or a decrease forecast. Moreover, we do not know at the beginning of a period whether positive or negative price changes will dominate that month. Therefore, the information vector must contain an average of the relative information gain and loss for each filter size. For a filter size of $X\%$, we would have

$$h(q) = \left[\frac{h(P_{0\cdot1}) - h\left(\frac{P_{11}}{P_1}\right)}{h(P_{0\cdot1})} + \frac{h(P_{0\cdot3}) - h\left(\frac{P_{33}}{P_3}\right)}{h(P_{0\cdot3})}\right] \Big/ 2. \tag{10}$$

This form was chosen over other possible averaging combinations, such as a weighted average, since it is both straightforward and since it allows us to

continue to say that the closer $h(q)$ is to 1·0, the better the forecasting ability of the filter size.

In an initial application of adaptive filtering models, the adaptive model (which we shall call adaptive model A) chose a filter size for next period on the basis of profitability and information gain. That is, the adaptive model employs as the optimal filtering size for period T that filter size which recorded the highest profits in $T-1$ subject to some minimum value of $h(q)$.[9]

A more simplified model called the pure information model was also employed because it too allows the filter size in a particular period to reflect new information. The model, however, does not contain a decision-maker environmental vector; it only has an historical information vector. The pure information model chooses as its filter size in period T that filter size in period $T-1$ with the largest cumulative information gain.

4. EMPIRICAL RESULTS

This model was run on two data sets September soybeans and September wheat for the period 1956–1970 with minimum information gain levels of 0·5, 0·6, and 0·7. In addition, a pure information gain model was employed which chose for the filter size in period $T+1$ that filter size with the highest information gain in period T. The profitability of these models is reported in Table 2.

TABLE 2. SUMMARY RESULTS

	Adaptive model A			Pure information model	Buy and hold
	0·5	0·6	0·7		
September wheat					
Gross profit	1925·00	2362·50	793·75	2368·75	−3656·25
Net profit	1313·00	2120·50	647·75	2042·75	−3937·25
Number of trades	52	22	14	30	13
September soybeans					
Gross profit	−3281·25	−250·00	700·00	2575·00	4568·75
Net profit	−4823·25	−1174·00	88·00	2227·00	4284·75
Number of trades	132	82	52	30	13
All data					
Gross profit	−1356·25	2112·50	1493·75	4943·75	912·50
Net profit	−3510·25	946·50	735·75	4332·75	347·50
Number of trades	184	104	66	60	26

[9] For this study the period is 1 month. Information gain statistics were developed for twenty filter sizes starting with 0·005 and terminating with 0·1.

The results indicate that both the pure information model and certain levels of the adaptive model outperformed the buy-and-hold strategy. Moreover, if we accept the argument that the proper criterion is simply the generation of positive profits, we can reject the Martingale hypothesis of commodity price movements. This conclusion is especially true in light of the results produced by the pure information model; nevertheless, an examination of the net profits produced by various models raises questions—particularly two interrelated questions that require additional investigation. First, why would the pure information model outperform the theoretically more complete adaptive model? Second, how sensitive are the results of the adaptive model to changes in the minimum information constraint? Both questions can be answered by comparing the variability of the net profits with the variability of the information about price movements produced by each filter size.

Examining the results derived for adaptive model A, we find that for both data bases (i.e. the soybeans and wheat) the model is very sensitive to changes in the minimum information gain constraint. Moreover, for both within and between the wheat and soybean data, there is no consistent relationship between the constraint level and the net profits produced by the model. This lack of consistency is extremely critical since it precludes the general applicability of each model.[10] In addition, these results are difficult to reconcile with the good results produced by the pure information model.

Vol. 4
1976

This lack of consistency led us to investigate the stability of the determinants of D_T, specifically $I_{T-\tau}$ (the historical information vector) and V_T (the environmental vector of the decision maker). For the adaptive model, the information vector contains for each filter size the relative average information gain of an increase and a decrease forecast while V contains the net profits for each filter in the preceding period.

An examination of the net profits produced by the individual filter sizes revealed an extremely high degree of variability in month-to-month performance. The variability in the net profit vector was reemphasized when juxtaposed with the stability of the information gain in the historical information vector.

It was concluded that the poor performance of the adaptive model (model A) could be improved by respecifying V_T. Thus, V_T was redefined to contain the cumulative net profits produced by each filter size over each commodity contract year. The new adaptive model (model B) chose as its filter size in period $T+1$ that filter with the highest cumulative net profits for the year in period T subject to the constraint on information.

Comparing the performance of models A and B, we find the latter produced

[10] In fact, examining the results within each data set forces the retraction of an earlier claim that raising the level of the information gain constraint tended to produce higher net profits [1, p. 497]. This erroneous conclusion occurred because initially the results of the adaptive model were evaluated at relatively few constraint levels.

superior results approx 80 per cent of the time. So, respecifying V reduced its variability and substantially improved performance.

Model B, however, did not indicate any stability with respect to the information constraint, nor did it produce a performance level equivalent to the pure information selection model. Again, the reason lies in the variability of the profit stream.

Table 3 reports the coefficient of variation of net profits and relative information gain for each filter size and commodity. The contrast is extreme. The adaptive model (either A or B) is not stable with respect to the information gain

TABLE 3. VARIABILITY COMPARISONS

	September wheat 1958–1970 coefficient of variation		September soybeans 1958–1970 coefficient of variation	
	Net profits	Relative information gain	Net profits	Relative information gain
0·005	−0·468	0·709	−0·481	−4·886
0·010	−1·210	0·080	−1·000	0·059
0·015	−1·542	0·067	−1·982	0·104
0·020	−1·795	0·103	−3·770	0·086
0·025	−9·236	0·180	−3·631	0·086
0·030	4·329	0·372	−3·446	0·033
0·035	3·155	0·709	19·941	0·119
0·040	10·630	1·174	8·162	0·127
0·045	30·083	0·191	3·898	0·115
0·050	−4·973	0·532	4·372	0·122
0·055	−9·049	0·163	2·814	0·040
0·060	−10·915	0·371	3·121	0·070
0·065	−19·630	0·210	4·277	0·148
0·070	5·961	0·219	5·451	0·177
0·075	11·661	0·209	4·632	0·178
0·080	17·892	0·384	4·608	0·224
0·085	3·238	0·935	5·386	0·482
0·090	4·062	0·948	7·522	−1·548
0·095	1·756	0·140	4·493	−1·798
0·100	1·649	0·118	4·959	0·000

constraint. And the reason for the lack of stability is the variability of the net profit stream produced by the individual filter sizes and used by the adaptive model in selecting next period's filter size. We are left with the conclusion that the underlying profitability of the process is no help in developing a profitable adaptive model. However, the forecasting ability of the individual filter sizes is sufficient by itself to produce consistent net profits. The reason for this apparent anomaly is simply that the variability of net profits produced by an individual filter size is much greater than the variance in the forecasting ability

of that same filter size as measured by information gain. Thus, we can conclude that individual filter sizes can be used to successfully predict commodity price movements even though there is no apparent relationship between predictive ability and short-term net profits.

One last observation needs to be stated. It was expected that the information gain would be useful in developing an adaptive model. Chi Square tests of independence on the prediction–realization tables (Table 1) consistently rejected the hypothesis that the predictions and realizations were independent. So, given the consistency of the information generated, its ability to choose superior filter sizes is not surprising.

5. SUMMARY AND CONCLUSION

In this paper the specification of the criterion for the evaluation of the efficient market hypothesis as it applies to commodity futures market was re-examined. It was established that all a filter model had to produce was positive profits, not profits greater than the buy-and-hold.

We then reexamined and extended the application of adaptive models to commodity futures and found that any adaptive model that directly considers the profitability of individual filter sizes does not produce superior results *vis-à-vis* a model that limits its concern to the information generated by the predictive power of the filter sizes. This result was due to the stability of the information gain statistic and the instability of the profitability stream.

In spite of our inability to develop a model that incorporates both information and profitability into the decision function, we should not forget that the pure information model run over a 26-year data base produced very profitable results—results that allow us to reject the Martingale hypothesis as it applies to the commodity futures market for wheat and soybeans.

REFERENCES

1. Martell TF and Philippatos GC (1974) Adaptation, information and dependence in commodity markets. *J. Fin.* **29**, 493–498.
2. Alexander SS (1961) Price movements in speculative markets: trends or random walks. *Ind. Mgmt Rev.* **2**, 7–26.
3. Jensen MC (1969) Risk, the pricing of capital assets, and the evaluation of investment portfolios. *J. Bus.* **42**, 167–247.
4. Leuthold RM (1972) Random walk and price trends: the live cattle futures market. *J. Fin.* **27**, 879–889.
5. Stevenson RA and Bear RM (1970) Commodity futures: trends or random walks. *J. Fin.* **25**, 65–81.
6. Murphy RE, Jr (1965) *Adaptive Processes in Economic Systems.* Academic Press, New York.

7. MARTELL TF (1973) Adaptive control models and the Martingale hypothesis. Unpublished PhD dissertation, The Pennsylvania State University.
8. THEIL H (1967) *Economics and Information Theory*. North Holland, Amsterdam.

ADDRESS FOR CORRESPONDENCE: *Professor TF Martell, The University of Alabama, College of Commerce and Business Administration, Alabama 35486, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

The Price–Quality Relationship: A Heuristic Model[1]

BM WAINSTEIN[2]

HS SICHEL[2]

(*Received November 1975; in revised form April 1976*)

An on-the-shop-floor experiment is carried out to determine the price–quality relationship for a number of toiletries. Three different and distinct demand curves are obtained. A heuristic model to describe these curves is constructed. The implications of this model for the profit-orientated businessman are then analysed.

Vol. 4
1976

INTRODUCTION

BUSINESS research is increasingly being directed towards the vital area of pricing. In particular, the relationship between price and imputed quality has come under the microscope of the applied economist. This paper deals with this theme and attempts to gauge the extent to which price is regarded as an authentic indicator of quality.

The pioneering work in this field appears to be that of Leavitt in 1954 [8]. He constructed a hypothetical situation whereby each respondent was confronted with two unnamed product brands and supplied with corresponding sets of prices. With price supplied as the only product differential, the respondents were asked to make a purchase decision. After making this decision, they were asked to answer questions concerning the satisfaction of their choices and to rate the products in regard to any perceived quality that they felt existed.

Leavitt's findings suggested that demand curves may not invariably be negatively sloped, that price itself may have more than one meaning to a consumer and that a higher price may sometimes increase rather than decrease his/her readiness to buy.

[1] Sections of this paper form part of an unpublished Master's thesis at the GSBA, University of the Witwatersrand [16].

[2] Barry Wainstein is currently employed as the Banking Editor of the Financial Mail. Herbert Sichel is Professor of Operations Research and Statistics at the Graduate School of Business Administration, University of the Witwatersrand.

This relationship was investigated from a different angle by three independent researchers in France. In 1954, Stoetzel postulated that consumers do not enter the market place with a fixed price for a particular product in mind, but rather with a range of acceptable prices. They refrain from buying below a certain limit, because they distrust the quality of the product, and above another limit because of expense. After experimentation, his overall conclusion was, ". . . since the individual frequently finds it impossible to know the net price of a good product being offered, he derives his notion of the value of an article from the price set by the vendor" [14, p. 74].

This work was further developed, firstly by Adam [1] and later by Fouilhé [4]. Although the results of these three studies are questionable because of inherent weaknesses in the experimental designs, they indicated that the concept of price as an indicator of quality warranted further attention.

Tull, Boring and Gonsior [15] conducted an experiment similar to that of Leavitt in 1964, with an added refinement. The key difference was that subjects were provided with a reference price, namely the price of the brand 'you normally buy'. It was felt that the addition of a reference price equalling that of the lowest-priced hypothetical brand provided would give each subject a similar reference point for his purchase decisions. If considerable purchase differences still occurred between 'essentially similar' and 'varied substantially' products, they believed that the hypothesis of price acting as an indicator of quality would be confirmed.

A definitive study was undertaken in 1966 by Gabor and Granger at the University of Nottingham [5]. The study was detailed and mathematically rigorous, based on a questionnaire which was submitted to 3191 respondents. Each respondent was quizzed on his upper price limit (the point above which he felt an article to be too expensive) and his lower price limit, (the point at which he might suspect the quality of an article). Although they concluded that the reputation of the manufacturer, the brand and the shop are important variables affecting consumer behaviour they could not deny that a reputation for high quality and high price do go together.

A number of studies were subsequently undertaken which involved more sophisticated tests than questionnaires.

To ascertain the attractiveness of a product, Ölander [10] presented products (towels) in a pairwise manner. Furthermore, because of a belief that the judgements would be better if some tangible benefit for the subjects arose, the subjects were allowed to retain their choice of product. He found that the ". . . results confirmed the hypothesis that towels are more often preferred when they are assigned a high price than when they are assigned a low price . . ." [10, p. 58].

McConnell [9] asked his subjects to select one of three brands of beer three times a week for 8 weeks. (They were actually the same product except for brand name and price.) Summarising his findings, McConnell states: ". . . price was used (by the subjects) as a cue to product quality. With a physically homo-

geneous product and unknown brand names (which have so little meaning that many never used them), the highest priced brand was perceived as being a better quality product than the other two brands" [9, p. 334].

In 1969, Stafford and Enis [13] attempted to clarify further the product–quality perception problem by extending research so as to include an independent variable other than price. They investigated the effect of store name on product–quality judgements, both separately and jointly with price. Their results suggested that consumer perception of price can depend upon other factors such as store prestige.

The above studies constitute the major undertaking in this field. Generally, the conclusions reached point to a direct relationship between price and quality for certain types of products. Before analysing these studies, it is necessary to look at the implications of their conclusions.

INTERPRETATIONS AND IMPLICATIONS

Vol. 4
1976

Shapiro has attempted to analyse the use of price as an indicator of quality and to explain the logic behind this perception. He claims that there are four reasons for consumers considering price to have this powerful piece of information.

(1) *Ease of measurement*
"Price is a concrete, measurable variable for a shopper. In most retail outlets price is fixed—not subject to bargaining. . . . Since price is concrete and measurable, the consumer views it with much confidence. He trusts it more than most cues directly concerned with quality" [12, pp. 21–22].

(2) *Effort and satisfaction*
Cardoza stated that "The effort invested in shopping may, under specifiable conditions, contribute to the evaluation of the product" [2, p. 248]. Shapiro maintains that the expenditure of money can often be equated to that of effort and subsequently the amount of satisfaction a customer receives from purchasing a particular product is directly related to the price of that product.

(3) *Snob appeal*
The Snob or Veblen effect is seen as a further reason for the importance of price. Although a person may know that the more expensive article is no better than the cheaper one, he may purchase it for prestige and social reasons.

(4) *Perceptions of risk*
"The prospective buyer balances
(a) the dollars-and-cents amount of the extra cost of a higher priced product against

(b) the possibility of losing out because of the assumed lower quality of the lower priced product. . . .

To reduce the risk of choosing a product of significantly poorer quality, the consumer chooses the higher priced brand" [12, p. 24].

This relationship, which has perceptively been interpreted by Shapiro, is summed up by Rabinowitz [11], who claims that price may be viewed as both an actual and a psychological indicator of quality. It is an actual indicator when the manufacturer sets prices on a 'cost-plus' basis and the customer, basing his perception on experience, acknowledges that the price is higher because costs are higher due to better quality materials and workmanship. The psychological aspect arises when quality differences among alternative products are seen to be great but cannot easily be judged.

The implications of these results should now be looked at in two ways.

Firstly, the postulation by the classical economists that demand curves generally slope downwards from left to right, thereby implying that as the price of a product increases the quantity demanded decreases, and vice versa, can be called into question. Thus, for certain product groups, classical demand theory, because of the tenuity of certain assumptions, is invalid.

Secondly, the concept of price as an indicator of quality has wide implications for marketers. If used discriminately, increased volume can be obtained by raising, not lowering, prices. Moreover, even if volume is not increased, revenue, the product of price and volume, may well be increased. In fact, both Knauth [7] and Dean [3] give examples of just such occurrences.

A CRITIQUE OF PREVIOUS METHODS OF INVESTIGATION

The consuming public is always the final evaluator of any product. Recognising this truth, Whitman [17] developed in-store experimentation and studied the effect on sales of price changes, without trying to keep "other things constant". He established "most precise and fruitful relationships" [17, p. 221] which summed up customer reaction to a market situation.

The experiment undertaken in this study is based on Whitman's method—but with certain modifications. Whereas he aimed at establishing a relationship between sales and price (and claimed "that there is no question that the response of sales to price is an elastic one" [17, p. 215] this study attempts to verify the hypothesis that price imputes quality. Every attempt is therefore made to keep other things equal and obtain a direct price–sales function.

By transposing Whitman's methodology—albeit modified—to the price–quality arena, a form of experimentation substantially different from those cited above is introduced.

The studies described earlier have all approached the problem hypothetically. Without exception, these studies have been based on the interview method. Although a few such as Ölander's [10] and McConnell's [9] approximate reality in the sense that the customer bases his simulated decision upon his perception of the product itself, rather than upon a generic description of it, the consumer, at no stage, bears any responsibility for an incorrect decision and is subsequently not motivated to indicate his true perceptions.

In particular, much of our knowledge of the price–quality relationship is based upon the authoritative study by Gabor and Granger. Moreover, the bell-shaped demand curve which is analysed later, is derived in this experiment [5].

It appears as if Gabor *et al.* are entirely satisfied with these results and no subsequent effort is made to improve upon the methodology. Instead an experiment was conducted whereby the hypothetical shop situation was tested for consistency. Their main conclusion (was) that the hypothetical shop situation method, being flexible, quick and cheap, can provide data of great reliability where there is a high degree of effective competition between brands. Where this does not apply, the hypothetical method can still be of considerable use, but the results will call for appropriate interpretation [6, p. 358].

Vol. 4
1976

In their, and other, surveys each subject was asked to respond to a number of prices called out to her by answering either "yes, buy" or "no, too expensive", and "no, too cheap". Questions of this nature are leading in that they supply motives to the respondents for justifying their actions. Hence, a bias is introduced into the question. It must be remembered that a common problem found in most market research is the respondent's desire to be as helpful as possible and to give answers which he believes will satisfy the interviewer. When leading questions are put, it becomes much easier to do this, thereby vitiating the results obtained.

By conducting an experiment on the shop-floor, this problem is avoided. Customers are able to indicate their attitudes to products by utilising their expendable funds. The actual purchase of a product, at a specific price, is, in the opinion of the writers, the only conclusive answer to the question we have been discussing.

Furthermore, an on-the-shop-floor experiment provides information about the real situation. Although Gabor's study was rigorous in that a large number of housewives were interviewed (unlike other studies, where the results were based upon the replies of a limited number of students) the situation was still a hypothetical one.

The results of the research study described in this paper show clearly how different the two methods are. Whilst Gabor had no difficulty in obtaining responses from approximately three thousand people, substantially less responses were obtained at the stores where this experiment was conducted.

On certain new products, which had not had sufficient opportunity to pene-

trate the market, it was not uncommon to find sales of under ten, even though two thousand customers had seen the products at specific prices. In contradistinction, had the same two thousand persons been interviewed, it is very likely that more than ten would have responded "Yes, buy". This point illustrates the difference in results when consumers are forced to back up their attitudes with committed actions.

A further criticism which may justifiably be levelled at most of the research mentioned in the literature survey is the inapplicability of the results. For example, McConnell's cognitive dissonance theory [9], which is elegant from the behavioural psychologist's point of view, provides no decision rules for the practical price maker. Even when recommendations to assist the businessman are put forward, these are limited in that they discuss selling prices without any reference to costs. Hence, the businessman is given no indication of profitability.

In deriving their bell-shaped demand curve, Gabor and Granger found a maximum response of 80 per cent, for most products. Similar results were obtained by Rabinowitz [11].

It is interesting to consider how valid a tool these demand curves truly are. The above experiments lead one to conclude that for any product, and, in particular, new products, the primary problem faced by the merchandising professional is the selection of the correct price. Although it has never been claimed that, at this price, the whole 80 per cent of the population will purchase the product, the discrepancy between the high response obtained in these experiments and ours nevertheless bears noting. Our results illustrate how difficult it is for a company to sell large quantities of any new product without a nation-wide advertising campaign and other promotional techniques. Even at a very low price, which is undeniably unfavourable from the company's point of view, fewer than 1 per cent of the customers making purchases buy the particular product.

In the Gabor–Granger type studies, where the 'response' is always quantal, in as much as the ordinate is the proportion of respondents who acknowledge that they will buy at specific prices, an added complication arises. Even if the quantal responses were realistic, it would still be impossible to convert the results into practical terms. "Yes, buy" does not indicate the number of units a person intends buying, nor does it specify the time period over which the purchases will be made.

In contrast, the experiment in this study was designed to allow for a down-to-earth interpretation in money terms. The response under index I, which is described shortly, indicates the number of units bought per number of customers exposed. This index has the advantage of being time-invariant, as a person is recognised to be another customer when that person returns to the store on another occasion.

EXPERIMENTAL DESIGN

In order to test the relationship whereby price imputes quality, an on-the-shop-floor experiment was deemed necessary [16]. A chain of discount stores, specialising in pharmaceuticals and toiletries, agreed to the running of such an experiment at two of its branches. The branches chosen enjoyed custom from persons in the higher income bracket, were relatively free from disruptive competition, and manifested a favourable image.

Prices on carefully selected products were manipulated at these branches. They were changed, on average, every fortnight in a random sequence to allow both high and low prices to be displayed at the beginning/end and the middle of the month. The number of units sold at each price was then recorded. From these results, attempts were made to derive demand curves.

It must be pointed out, however, that experimentation with prices below cost was limited. It has been suggested that if the price range is appropriately extended the positively sloped part of the curve will always appear, unless customers are convinced that the product quality has not deteriorated at the lower prices. This applies even if a plateau and not a peak occupies the centre. Practical considerations of profit precluded the investigation of this hypothesis.

Vol. 4
1976

Although the experiment was designed to run over fortnightly periods, the following problems arose:

(1) Public holidays affected the number of trading days.

(2) This problem was compounded because certain of these holidays fell on weekends (busy periods) whilst others fell during the week (quieter periods).

(3) A month-end falling during a period had a considerable effect upon results, due to the increased customer traffic.

(4) Since certain observations were disqualified due to stockouts, promotions, miscounts, etc., it became necessary to expedite the experiment by shortening some time periods.

The crucial factor affecting the number of units purchased is the number of customers entering the store. Hence, to normalise the periods, the following index was used:

$$I(10^k) = \frac{N}{R} \times 10^k,$$

where $I(10^k)$ is the adjusted quantity sold per 10^k customers; N is the number of recorded units sold during a particular period, at a particular price; R is the number of customers making any purchase during that particular period; and k is a power of 10.

A pilot study run earlier indicated that certain toiletries were suitable for experimentation. Nine different products were chosen, of which six are described

below. Problems of time, cost and administration precluded a larger number of items, but it was hoped that in spite of its smallness the experiment would illustrate the relationships hypothesised. Indeed, despite the paucity of the data the results clearly throw light on the subject.

The products chosen were the following:

(1) *Hudson Rainbow Tissues*
This product is a paper tissue, a category which shows high market sensitivity. In other words, the volume of purchases is demonstrably affected by price changes.

(2) *Femfresh Deodorant*
This is a lady's deodorant which was new on the market at the time the experiment was initiated. It was selected because Leavitt's study [8] indicated that deodorants fell into that category of products in which consumers believed that they could perceive quality differences.

(3) *Lorde Blades*
Leavitt [8] also found that razor blades were perceived by customers to be considerably different. On this basis, a recently launched high quality blade was chosen.

(4) *Clinic Active Shampoo*
Tull *et al.* [15] found in their research that customers believed different brands of liquid shampoo to be qualitatively different. Thus, this high price, high quality shampoo was chosen.

(5) *Contra Shampoo*
This again is a quality anti-dandruff shampoo, publicised and recognised as such. The test on this item was run in conjunction with that on the previous item, in order to gauge the differences in behaviour, if any. Certain differences in advertising and image exist between the two products.

(6) *Brylcreem Natural Control*
A men's hairspray which had, shortly before the inception of the study, been introduced into a highly competitive market was also selected.

The products described above were displayed—in the same positions for the duration of the experiment—on a prime-end. Since the chosen end was the most favourable in the store and was that end that the largest number of customers were forced to pass, it was possible to obtain maximum exposure for these products. Furthermore, since the total number of customers making a purchase, per day, was known, it was reasonable to assume that a high percentage of this same number of customers passed the products. (A test had earlier been conducted to ascertain the high correlation between the number of customers entering the store and those making purchases.)

In addition to this display, the products were also exhibited in their normal positions, among their competitive brands. Hence, if customers wanted a particular product, they could obtain it where they customarily found it. Depending upon the price, they would take or leave the product. The display on the prime-end, on the other hand, was designed more for the impulse purchase—the main objective. On seeing the product, the customer, judging the value primarily by price, would or would not make a purchase.

In determining the motivation behind consumer behaviour, a number of elements of the marketing mix, other than price and/or quality, come into play. Every effort was made in the design of this experiment to ensure that conditions in the branches remained as stable as possible during the experimental period.

(1) In-store promotion was neutralised in the sense that the same prime-ends, shelf-space allocation and price tags were used throughout. No attempt was made to promote any item unduly. When certain of the products were included in a weekly promotion, that week was discounted and the results omitted.

(2) No advertising on the experimental items was undertaken. Furthermore, if competitors strongly advertised a certain line, which would draw customers into that store rather than a branch of the sponsors, for that particular item, the effect was noted. The same procedure applied when the sponsors advertised a rival brand different from the one on display.

(3) No record was kept of prices of these products at the competitive stores. Firstly, the branches in which the experiment was run were chosen because they were faced with minimal competition. Secondly, the task of monitoring prices on nine products at the large number of outlets in Johannesburg is prohibitive and hence competitors' prices were regarded as a normal hazard of business.

However, it must be admitted that in an experiment of this nature, certain undesirable influences will persist. As will be seen, the results obtained were generally not unduly biased by these extraneous factors, although substantial scatter of the data is evident.

RESULTS AND A HEURISTIC MODEL

The actual results of the experiment on the shop-floor are given in Tables 1–3 below. The first column in these tables gives the sales price at which the products were displayed and sold.

TABLE 1. ELASTIC DEMAND CURVES

HUDSON RAINBOW TISSUES

Price x (cents)	Standardised sales $I(10^3)$	Number of units Sold (N)	Number of purchasing customers (R)
18	23·31	216	9268
18·5	35·84	288	8035
19	15·17	130	8571
20	16·26	158	9719
21	14·57	252	17,297
22	11·87	216	18,196
23	9·30	180	19,358
24	8·07	144	17,840

FEMFRESH DEODORANT

Price x (cents)	Standardised sales $I(10^4)$	Number of units sold (N)	Number of purchasing customers (R)
72	30·66	49	15,980
83	30·51	61	19,993
92	8·40	19	22,617
101	15·90	26	16,354
110	6·97	17	24,390
119	4·25	10	23,528
128	11·97	32	26,738

TABLE 2. BELL-SHAPED DEMAND CURVES

LORDE BLADES

Price x (cents)	Standardised sales $I(10^4)$	Number of units sold (N)	Number of purchasing customers (R)
27	2·17	2	9233
33	7·15	13	18,196
35	4·92	9	18,296
37	18·50	33	17,840
39	9·89	20	20,219
41	1·74	3	17,273
43	7·52	13	17,297
45	3·17	6	18,953
47	0·00	0	19,353

CLINIC ACTIVE SHAMPOO

Price x (cents)	Standardised sales $I(10^4)$	Number of units sold (N)	Number of purchasing customers (R)
73	1·06	1	9470
78	3·92	7	17,835
83	5·69	6	10,536
88	3·85	7	18,178
93	1·78	3	16,846
98	14·01	27	19,272
103	5·31	10	18,844
108	14·00	26	18,570
113	0·00	0	9553

TABLE 3. INELASTIC DEMAND CURVES

CONTRA SHAMPOO

Price x (cents)	Standardised Sales $I(10^4)$	Number fo units sold (N)	Number of purchasing customers (R)
73	10·56	10	9470
78	16·26	29	17,835
83	10·44	11	10,536
88	14·30	26	18,178
93	8·31	14	16,846
98	19·20	37	19,272
103	9·55	18	18,844
108	4·31	8	18,570
113	6·28	6	9553

BRYLCREEM NATURAL CONTROL

Price x (cents)	Standardised sales $I(10^4)$	Number of units sold (N)	Number of purchasing customers (R)
63	7·65	14	18,296
67	12·16	21	17,273
71	9·34	17	18,196
75	9·30	18	19,358
79	5·44	11	20,219
83	8·37	8	9553
87	10·41	18	17,297
91	5·04	9	17,840
95	0·00	0	18,953

Vol. 4
1976

The second column shows the standardised sales, i.e. the number of units sold per 10^k purchasing customers at the store.

The third column indicates the actual number of units sold. With the exception of the Hudson Rainbow Tissues in Table 1, it will be noticed that the number of items purchased is quite small. One must therefore expect that the sampling variations of the observed frequencies of sales, for a given sales price, should approximately follow the Poisson distribution law as the 'exposure', i.e. the number of purchasing customers R, shown in the last column of the tables, is fairly large.

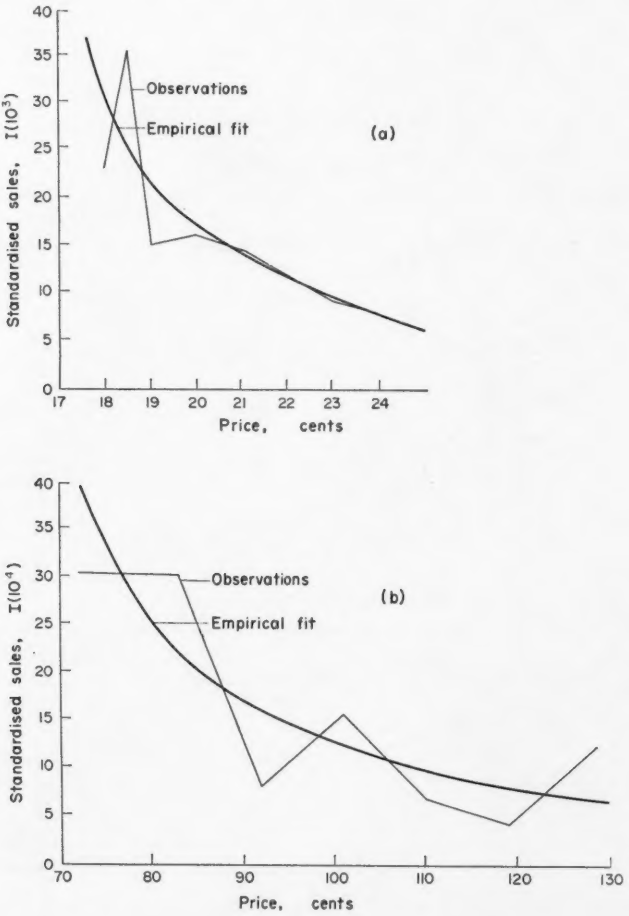


FIG. 1a. *Hudson Rainbow Tissues.* 1b. *Femfresh Deodorant.*

Vol. 4
1976

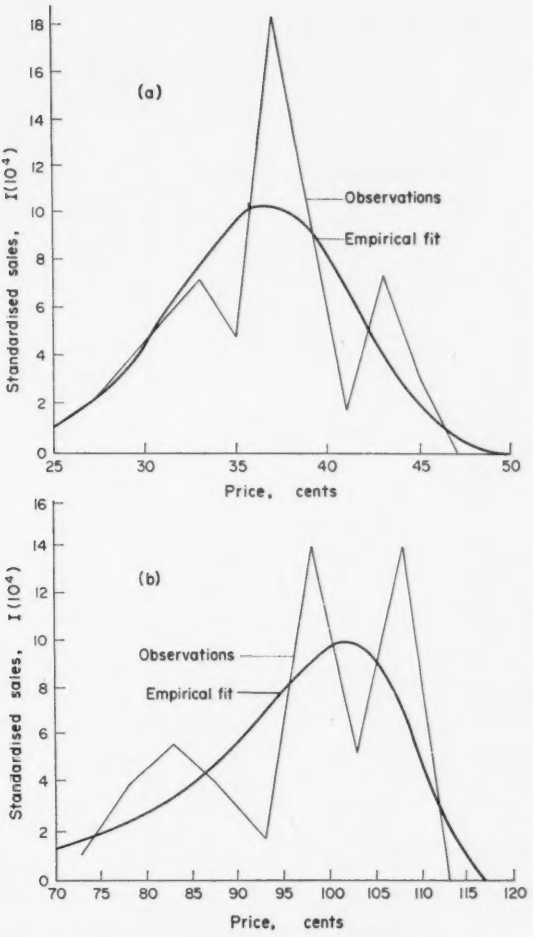


FIG. 2a. *Lorde Blades.* 2b. *Clinic Active Shampoo.*

The relatively large sampling variations of Poisson variables with small expectations can be noticed in Figs. 1–3 where the observed standardised sales were plotted against the unit price.

However, the shape of the actual 'demand curves' is beyond any doubt in spite of the individual fluctuations around the regression graphs.

In all, three types of demand curves were observed during this investigation. They are shown as:

Figures 1a and b indicating classic negative downward sloping curves which, for short, we call 'elastic'.

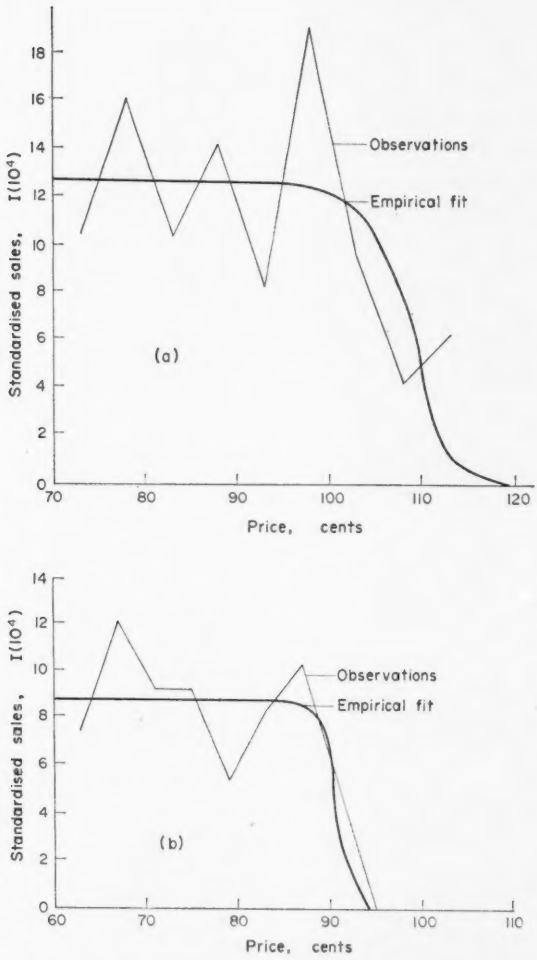


FIG. 3a. *Contra Shampoo.* 3b. *Brylcreem Natural Control.*

Figures 2a and b display sales resistance at high *and* low prices in line with the price–quality relationship discussed by Gabor and Granger [5]. This demand curve we call 'bell-shaped'.

Finally, a rather unusual type was encountered and is shown in Figs. 3a and b. Here the sales resistance is, as always, at the high end of the price range, whereas the standardised sales remain at a constant level as we proceed from high to low prices. This phenomenon we call 'inelastic'.

TABLE 4. PARAMETER ESTIMATES FOR REGRESSION MODEL

Figure	Product	Price range (in cents)	Cost price 'c' (in cents)	Parameter estimates $\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	Type of demand curve
1a	Tissues	18–24	17	−0·1903	$1·2057 \times 10^{10}$	−1·3755	778·0	Elastic
1b	Deodorant	72–128	75	−0·01776	79·986	−0·06430	66·50	Elastic
2a	Razor blades	27–47	29	0·6844	−1·2884	0·05951	$1·2030 \times 10^{-5}$	Bell-shape
2b	Clinic shampoo	73–113	71	0·07941	$-0·4612 \times 10^{-7}$	0·1588	$0·4955 \times 10^{-2}$	Bell-shape
3a	Shampoo	73–113	71·5	0	$-4·0012 \times 10^{-16}$	0·3224	12·80	Inelastic
3b	Haircream	63–95	61	0	$-4·3862 \times 10^{-32}$	0·7879	8·940	Inelastic

Although satisfactory models—particularly that of Gabor and Granger [5] describing the bell-shaped curve—have previously been formulated, it is desirable to have one single regression model for the three types of demand curves found in this investigation. We propose the following function

$$f(x) = \partial e^{\alpha x + \beta e^{\gamma x}}, \tag{1}$$

where $f(x)$ represents the number of standardised sales as a function of unit price x. (It will be recalled that the standardised sales $I(10^k)$ are independent of the time period in which sales are made. They are dependent only on the number of units purchased during *any* period and the number of purchasing customers in the store during *that* same period.)

In the case of the elastic demand curves, parameters α and γ are negative and β is positive.

Bell-shaped demand functions arise if parameters α and γ are positive and β is negative.

The inelastic model is obtained if parameter $\alpha = 0$, β is negative and γ is positive.

Parameter estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ are tabulated (Table 4) for the six products dealt with in this investigation together with the price ranges over which the experiments were carried out. In addition, the cost prices to the store, for all items, are also listed.

The resulting six regression curves have been plotted in Figs. 1a–3b.

For the bell-shaped demand functions, the maximum standardised sales are obtained, at a price x_{max}, from

$$x_{max} = \frac{1}{\gamma} \ln \left(\frac{\alpha}{|\beta| \gamma} \right), \tag{2}$$

where β is in modulus form.

For example, the unit price for which sales are optimized may be obtained by substituting parameter estimates from Table 4 into equation (2).

Thus we found that for Lorde Blades (Fig. 2a)

$$x_{max} = 37 \text{ cents}$$

and for Clinic Active Shampoo (Fig. 2b)

$$x_{max} = 102 \text{ cents.}$$

If x is the unit sales price and c is the unit cost price to the store, the total standardised profit will be

$$P(x) = (x - c)f(x). \tag{3}$$

Maximising the total profit in (3) gives

$$f(z_{max}) = -(z_{max} - c)f'(z_{max}) \tag{4}$$

where $f'(z_{max})$ is the first derivative of the standardised sales $f(x)$ at the profit optimum z_{max}.

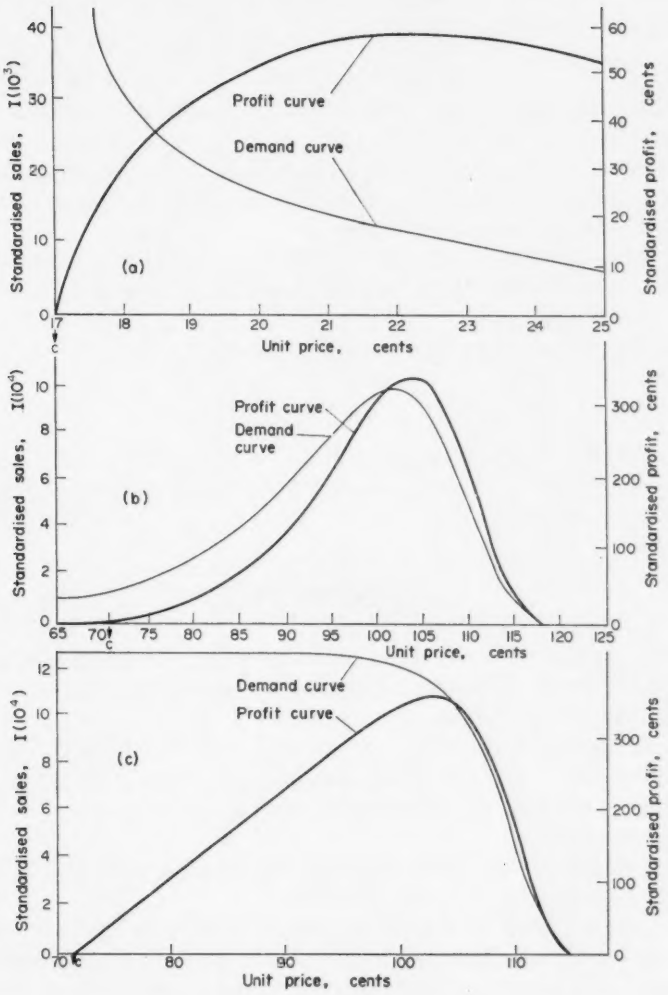


FIG. 4a. *Hudson Rainbow Tissues.* 4b. *Clinic Active Shampoo.* 4c. *Contra Shampoo.*

The RHS of equation (4) must be positive as the LHS is always positive. As $(z_{max}-c)$ must be positive for achieving a profit, it follows that $f'(z_{max})$ must be negative.

Hence z_{max}, the unit price for which the total profit is maximised, must lie to the right of the asymptote in the case of the inelastic demand curve.

Vol. 4
1976

For the bell-shaped response curve, the profit maximum will lie to the right of the sales maximum, i.e.

$$z_{max} > x_{max}.$$

This means that the optimum profit is not at the sales maximum but slightly above it.

In the case of the elastic demand curve, one has to investigate carefully whether a profit maximum in fact exists. If it does, one will frequently find that it is a fairly flat optimum, as in Fig. 4a.

The various types of demand curves and their associated standardised total profit graphs are shown in Figs. 4a–c. All profit curves begin at the cost price c, where the total profit is nil.

By substituting equation (1) into (4) one obtains

$$e^{\gamma z_{max}} = -\frac{\alpha}{\beta\gamma} - \frac{1}{\beta\gamma(z_{max} - c)}, \tag{5}$$

If parameters δ, β and γ are replaced by their estimates in Table 4, it is possible to calculate z_{max} (the unit price at which total profits are maximised), by solving equation (5) by iteration.

DISCUSSION

As seen from Figs. 1–3, the demand curves for the products chosen are well described by the heuristic model. It is now necessary to give the rationale for the behaviour of these curves.

(1) *Elastic demand curves.* Both *paper tissues* and *deodorants* are product types which are reputed to show price elasticity due to an acute price consciousness by consumers. In fact, this observation was one of the primary causes for selecting these products.

(2) *Bell-shaped demand curves.* There are two possible explanations for the shape of the demand function of *Lorde blades.*

Firstly, at the time of the experiment, this brand of razor blades was *new* on the market —a category suggested in the literature to show this phenomenon.

Secondly, as pointed out by Leavitt [8], customers are quality conscious when purchasing blades, and consequently have more faith in the more expensive brands. In other words, only at a price which indicates that it is a quality blade will customers be induced to make purchases. However, sales drop off rapidly as the price becomes prohibitively expensive.

Clinic Shampoo is recognised to be a quality shampoo for which the postulates of the bell-shaped curve are applicable.

(3) *Inelastic demand curves.* An important observation which one can make when considering *Contra Shampoo* is that the response curve is significantly different from that of Clinic shampoo.

The reason for this lies in the dissimilar characteristics of the products. Contra is heavily promoted and advertised as an anti-dandruff *treatment*, not as an anti-dandruff shampoo. Thus, its image is that of a *medicinal* product, not of a personal hygiene product. If a customer believes in the efficacy of a medicine, she will purchase it whenever the need arises, irrespective of price. This product loyalty accounts for the stable sales volume over a wide range of prices.

The consumer who purchases *men's hairsprays* is usually the young sophisticated man, who recognises the attraction of the product in spite of its traditional feminine orientation. He is neither deterred by a high price nor induced to make a purchase by relative inexpensiveness.

This brief discussion has shown that each product conforms to a particular case of the heuristic model for reasons of consumer psychology, product image and perceived quality differences. By understanding the appeal of a product in these terms and offering it in the relevant market, the businessman can use this model to maximise his prime objective of profitability.

CONCLUSION

Vol. 4
1976

As we have defined three distinct demand curves, it is necessary to determine the conditions under which profits are maximised in each case.

Figure 4 illustrates the relationship between the demand curves and their associated profit graphs, profit being the product of demand and gross markup. Of course, the peak of the profit curve depends on the position of the cost price even when the demand curve, over a constant price range, is the same.

The diagrams indicate the sensitivity of the different models under an invariant cost price. A very pertinent consideration concerning sensitivity which the price-maker is entitled to ask is: "What are the consequences of an incorrect price?"

Figure 4a shows that for a product displaying elastic demand, the curve is more vulnerable on the downside. In other words, a price which is too low produces a lesser profit, in spite of greater demand, than one which is too high. In fact, we see that profit decreases slowly as the price increases, and that we experience a plateau.

This observation is different from the popular view of many marketers who believe that the way to increase profits is by expanding turnover, at the cost of a smaller margin. The view raised here not only contradicts that logic but also has other attractions. A smaller volume of sales requires a smaller inventory and hence allows for greater liquidity. For small businesses, with cash flow problems in particular, this aspect of the study should be most appealing.

Caution must, however, be exercised when adopting this strategy. The experimental prices ranged over a limited spectrum and any extrapolation far above the extreme price tested is subject to error. There is no empirical evidence that sales resistance does not develop rapidly beyond a certain price, and it is

possible that the tail of this function behaves similarly to that of the inelastic curve. Such a conclusion implies that the elastic and inelastic graphs discussed above are basically different parts of a single compound demand curve.

Figure 4b deals with profitability associated with a bell-shaped demand curve. Products evincing this curve require the most carefully analysed price structure, as considerable sensitivity is seen on both sides of the rather peaked maximum.

Finally, Fig. 4c deals with the profit sensitivity of an inelastic curve. It is evident that considerable upside risk exists. On the downside, however, one observes that profit is a function of price only, since demand is inelastic. The price-setter is thus faced with the situation whereby a 'too-high' price will inhibit sales altogether, whereas a cautious 'too-low' price affects only the profitability.

REFERENCES

1. ADAM D (1969) Consumer reactions to price. In *Pricing Strategy*. (Eds. TAYLOR B and WILLS G) pp. 75–88. Staples Press, London.
2. CARDOZA RN (1965) An experimental study of customer effort, expectation and satisfaction. *J. Marketing Res.* **2,** 244–249.
3. DEAN J (1951) *Managerial Economics*. Prentice-Hall, Englewood Cliffs, New Jersey.
4. FOUILHÉ P (1969) The subjective evaluation of price: methodical aspect. In *Pricing Strategy*. (Eds. TAYLOR B and WILLS G) pp. 89–97. Staples Press, London.
5. GABOR A and GRANGER CWJ (1966) Price as an indicator of quality. *Economica* **46,** 43–70.
6. GABOR A, GRANGER CWJ and SOWTER AP (1970) Real and hypothetical shop situations in market research. *J. Marketing Res.* **7,** 355–359.
7. KNAUTH O (1949) Considerations in the setting of retail prices. *J. Marketing* **14,** 1–12.
8. LEAVITT HJ (1954) A note on some experimental findings about the meanings of price. *J. Bus.* **27,** 205–220.
9. MCCONNELL JD (1968) Effect of pricing on perception of product quality. *J. Appl. Psychol.* **52,** 331–334.
10. ÖLANDER F (1969) The influence of price on the consumer's evaluation of products and purchases. In *Pricing Strategy*. (Eds. TAYLOR B and WILLS G) pp. 50–74. Staples Press, London.
11. RABINOWITZ P (1971) Formulating a pricing policy for pot o' gold products. Master's Thesis, Unpublished. University of the Witwatersrand.
12. SHAPIRO BP (1968) The psychology of pricing. *Harv. Bus. Rev.* **46,** 14–25, 160.
13. STAFFORD JE and ENIS BM (1969) The price–quality relationship: an extension. *J. Marketing Res.* **6,** 456–458.
14. STOETZEL J (1969) Psychological/sociological aspects of price. In *Pricing Strategy*. (Eds. TAYLOR B and WILLS G) pp. 70–74. Staples Press, London.
15. TULL DS, BORING RA and GONSIOR MH (1964) A note on the relationship of price and imputed quality. *J. Bus.* **37,** 186–191.
16. WAINSTEIN BM (1973) Some aspects of consumer behaviour with respect to the price–quality relationship. Master's Thesis, Unpublished. University of the Witwatersrand.
17. WHITMAN RH Demand functions for merchandise at retail. In *Studies in Mathematical Economics and Econometrics*. (Ed. LANGE O) pp. 208–221. University of Chicago Press, Chicago.

ADDRESS FOR CORRESPONDENCE: *BM Wainstein, 31 Jacmor Court, Cor Cavendish Road and St Georges St, Yeoville, Johannesburg, South Africa.*

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press .Printed in Great Britain

Assessing Operational Policies

E STEPHENSON

(*Received November 1975; in revised form March 1976*)

This paper demonstrates the value and use of continuous simulation as a means of assessing operational policies prior to their implementation. Using as an example the production and employment system of a firm producing specialist pumps to order, it is shown that such assessments are practical and more likely to yield valid results than the unaided and more intuitive assessment by management. Successful application of the method depends upon the full utilisation of management's knowledge of the system, but the simulation approach used has the advantage that it can be understood, relatively easily, by managers with little, if any, mathematical training.

Vol. 4
1976

1. INTRODUCTION

THIS paper illustrates the use of continuous simulation to investigate the effect on performance of particular policies adopted, or being considered for adoption, by a firm. Such investigations provide a means of comparing competing policies and hence assist in the selection of those most likely to achieve efficient performance. To avoid unnecessary complexity the case described is limited to the production and employment functions, but in principle the method may be used to investigate any area of a firm's operations or even the total operation of the firm.

Stated briefly, and at the risk of oversimplication, a manufacturing firm, in order to satisfy at a profit the demand for its products, operates within the production function certain policies with regard to, for example, labour input, production rate and inventory. Such policies aim at filling the maximum number of orders on time and at acceptable cost. Control of operations to achieve conformity with these policies is based on the measurement of key variables such as labour costs, production rate, deliveries, work in progress (inventory), order backlog etc. Such measurement is done via the firm's system of performance controls, the differences between target and actual values being both a measure of performance and the information on which decisions generating future action are based. Future action will be directed towards correcting divergences from the behaviour required by the agreed policies or toward pursuing new policies if current ones prove inadequate.

Assessment of the stated policies in terms of system performances, as measured

by the differences between target and actual values of key variables, is necessarily an after-the-event procedure, since obviously knowledge of performance must follow actual performance. However, once a set of policies is adopted, the type of procedure, briefly indicated above, must be followed if adequate control and satisfactory performance is to be maintained. The question arises as to whether it is possible to realistically assess the effects of a given set of policies before they are implemented. If, for example, the firm expects to experience conditions of demand which differ markedly from current conditions then responsible management will attempt to assess the effect upon performance of operating current or alternative policies under the changed demand conditions. Unfortunately, because of the usually complex interactions between the system variables and particularly since the prediction of these variables over time is involved, accurate assessment is beyond unaided human ability.

An individual or group of individuals attempting to carry out such an assessment essentially build a mental or verbal model of the system under consideration which incorporates the various material, men, money and information flows, the decision rules implied by the given policies, the mechanism by which decisions are translated into action and the time delays between receipt of information, implementation of decisions and the results of the ensuing actions. They then attempt to trace the effects upon the various parts of the system of a particular input signal, for example, the changed demand for the firm's products. They are in effect attempting to carry out a continuous simulation over time of the behaviour of the system and unaided they will either fail or will simplify the system to such an extent that their predictions will be of doubtful value. With the aid of a computer however, such a simulation can be successfully conducted; in the case described here, Forrester's method is used [1], the computations being carried out on an IBM 360/65 machine using the DYNAMO compiler [4]. Other continuous simulation programmes, for example, CSMP [2] are available.

The method requires a detailed knowledge of the structure of the system under consideration, of the information links between the various parts of the system, of the delays which exist between various points and of the decision rules which translate policies into actions on the basis of information received. Such knowledge, which can only be provided by management, enables the analyst to build a model of the system which can be represented both by a flow diagram, such as shown in Fig. 1, and by a set of relatively simple linear and/or non-linear equations of the type briefly illustrated in the next section. The computations essentially consist of making a change in some input signal, for example order rate, and calculating the resulting values of all the variables defined by the equations. This procedure is repeated at successive small time steps for the whole period of time the investigation is to cover and hence a continuous record over time is obtained of the behaviour of the system as represented by the values of the variables describing the system.

Vol. 4
1976

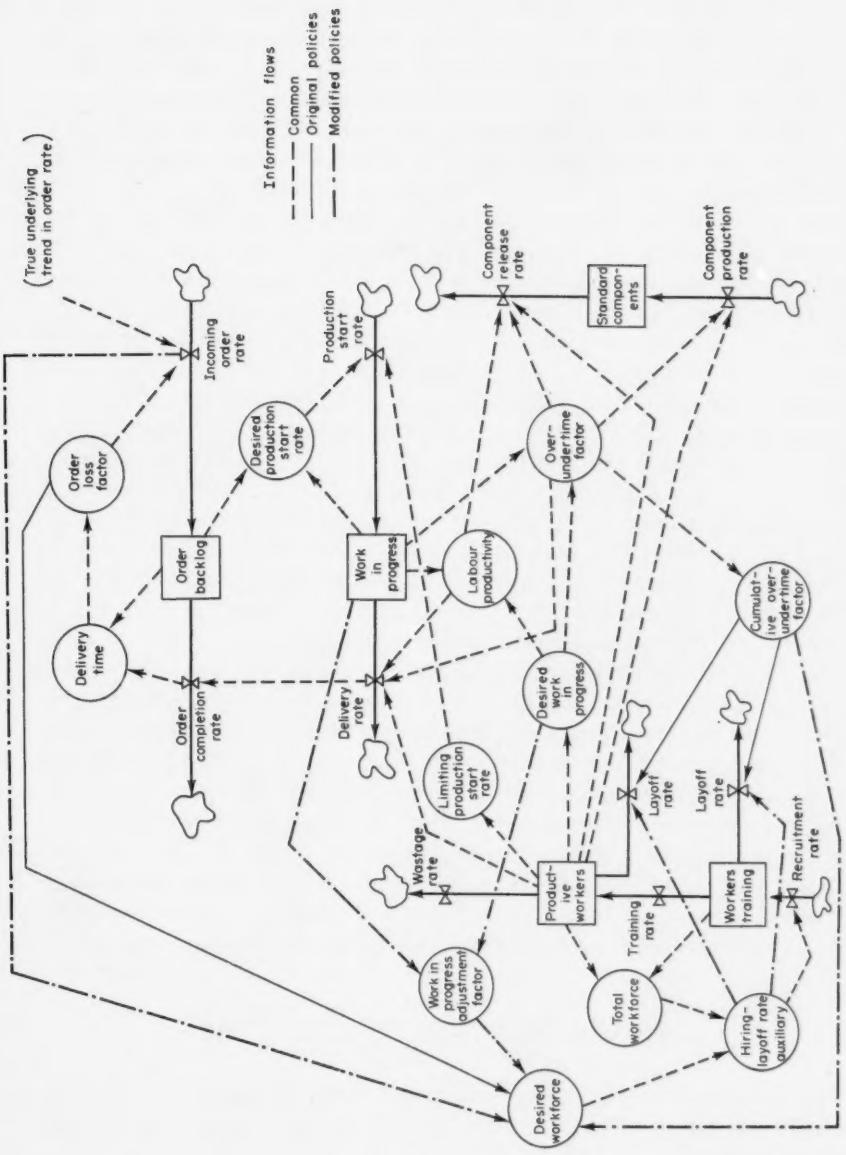


FIG. 1. *Simplified flow diagram.*

If the system behaviour is judged to be unfavourable, with reference to certain agreed performance criteria, the equations can be modified to represent alternative policies and the simulation re-run until acceptable behaviour is predicted. Obviously the method requires close co-operation between managers and analyst, since valid prediction is dependent upon a realistic description of the system which only management can provide. The method also has the advantage that the equations are written in a form which can be understood, relatively easily, by managers who do not have a mathematical training.

The method is illustrated in this paper by calculating the response of the production and employment system of a firm to a change in demand for its products. The firm manufactures, to order, a wide range of specialist pumps and has been operating under conditions in which the underlying trend in demand has been approximately constant. Under such conditions the workforce of 180 has maintained steady production such that the average time between receipt of orders and despatch of goods has remained essentially constant at 7 weeks. However because of recent technical improvements in the product, management expect that orders will commence to increase in about 6 months time and could rise, over a 6 month period, to a level possibly 50% higher than the current level provided a delivery time of 7 weeks can be maintained. In order also to examine the response of the system under conditions of falling demand, it is artificially assumed that after about a year at this higher demand level there is a step fall to the original level, at which level demand is maintained for the remainder of the investigation.

The possibility of a changing demand raises the question as to whether or not current policies with regard to the hire and layoff of labour, overtime, production rate, etc. will continue to result in adequate system performance. In particular, it has been suggested that the present policies relating to the hiring and firing, which essentially are simple reactions to problems as they arise, will prove inefficient and should be replaced by policies based on a forecast of demand. Proving this requires a comparison of simulation runs which incorporate the two alternative sets of policies; the basis on which the comparisons are made is discussed in a later section. The next section describes the system and the various policies adopted, such description being the basis for the system model.

2. DESCRIPTION OF THE SYSTEM

For some time the demand for the firm's products has, apart from random fluctuations, been essentially constant at 30 units per week and the present production workforce of 180 men working normal time and at normal productivity (0.167 units/man-month) has been able to satisfy the demand and maintain a delivery period which has varied little from the quoted value of 7 weeks. During periods of slack demand, normal working has been maintained by using

the workforce to produce standard components for stock which are released when demand recovers. Thus while stocks last production time for complete units is reduced.

The production rate achieved by the workforce is affected by two factors in addition to the standard component factor referred to above; these are the use of over/undertime and labour productivity. It is the firm's policy to release orders to the shop floor with minimum delay so that a change from normal demand requires the use of over/undertime for the production of complete units, the amount of over/undertime required being directly proportional to the amount by which actual work-in-progress differs from the desired level. The latter is that level for which, with the current workforce working normal time and at normal productivity, the production time per unit is 6 weeks. Overtime however is never allowed to exceed 30% of normal time in any period. In modelling the system, labour productivity was assumed to depend only upon the actual and desired levels of work-in-progress. When actual is less than or equal to the desired level, labour productivity is unchanged from its normal value, but as actual increases above the desired level, bottlenecks begin to develop on the shop floor and labour productivity decreases, the rate of decrease increasing the more actual exceeds desired level.

Vol. 4
1976

The development of bottlenecks and consequent decrease in productivity results in an increase in production time and this, together with the increased time for which orders may have to wait before release to production, increases the delivery time. There is a further delay before customers become aware of this increase but as the latter is recognised complaints begin to increase and, unless conditions improve, orders are lost. Thus the incoming order rate is influenced by delivery performance and although empirical data on this effect is difficult to obtain it is possible to make some reasonable estimate of the relationship.

The firm has no declared policy on recruitment but effectively the existing policy is based on the order loss factor referred to above. If under conditions of increased demand a loss of orders occurs, recruitment takes place at a rate such as to remove, over a period of 3 weeks, the discrepancy between actual and desired workforces. The latter reflects management reasoning that, approximately, an $x\%$ loss of orders is caused by an $x\%$ increase in work-in-progress above the desired level and therefore to restore production and delivery times to their normal values requires an $x\%$ increase in the existing workforce. New recruits undergo a 10 week training period. Under reduced demand conditions workers in training are laid off before fully trained workers, the total layoff rate being based on records kept of the cumulative undertime man hours on complete unit production. The equivalent overmanning during the previous 8 week period is determined and this number of men laid off over a 2 week period.

Under the influence of the various policies described, relating to production

rate, work-in-progress, over/undertime, component production and hiring and firing, and of the system driving signal which is the changing demand pattern described in the introduction and shown in Fig. 2, the system will exhibit certain behaviour and since the purpose of this study is to compare this behaviour with that if alternative policies were adopted, the latter are described next. In the investigation reported here, only the policies relating to hiring and firing were changed.

In discussing possible policy changes some of the management team criticised the present policy of delaying recruitment until customer complaints increased and argued that particularly adverse consequences would result under conditions of increasing demand. In the longer term such a policy must damage the firm's reputation and lead to the permanent loss of customers and in the short term would require the excessive use of overtime in attempting to satisfy demand. To avoid these undesirable consequences it was proposed that recuitment/layoff should be based on a forecast of the underlying long term trend in demand, with short term variations being absorbed by the use of over/undertime.

The recruitment/layoff decision rules adopted specified that the discrepancy between existing and desired workforce be removed over a period of 3 or 2 weeks according to whether recruitment or layoff was required. The desired workforce is determined by three factors, the principal one being the forecast of long term trend in demand. If hiring/firing is to reflect only long term changes in demand, the forecasting method used must suppress short term changes while revealing the long term trend; this was achieved by using the double smoothing method [3] with a relatively long smoothing constant. The corresponding component of the desired workforce is then determined by dividing the forecasted rate by the value of normal productivity.

The second factor is concerned with maintaining the actual level of work-in-progress at the desired level to avoid the development of bottlenecks on the shop floor. To achieve this the desired workforce includes the number of men required to remove the difference between the two levels over a period of 3 weeks. Finally to compensate for any over/under-manning in the short term, the desired workforce is adjusted by the men equivalent to the average, over the previous 3 weeks, of the cumulative total of over/undertime man hours. This latter adjustment is made only if, in the case of overmanning the workforce is currently on undertime or in the case of undermanning the workforce is currently on overtime.

The type of system description given above is the starting point for modelling the system and obviously the realism of the model is highly dependent upon the accuracy and completeness of that description, which implies the close involvement of management in the modelling process. The latter consists of transforming the verbal description into a set of equations suitable for solution by the particular simulation programme used. To illustrate, the following three

equations describe and determine the amount of work-in-progress at time K.

$$I.K = I.J + DT^* (P.JK - DR.JK); \quad (1)$$

$$P.KL = \min (DP.K, \text{LIMP}.K); \quad (2)$$

$$DR.KL = PW.K^* (I+\text{OUF}.K^* LP.K/\text{PDCF}.K \quad (3)$$

Equation 1 simply states that work-in-progress (I) at time K is equal to that at time J plus the change occurring in the interval DT between times J and K. This change is the result of the difference between the production start rate (P) which increases work-in-progress and the rate at which completed units are delivered (DR) which decreases I. Equation 2 defines P as the minimum of the desired production start rate (DP) and the limiting production start rate (LIMP) which reflects the 30% constraint on overtime. The delivery rate (DR), equation 3, is seen to depend upon the size of the productive workforce (PW), the fraction of over/undertime (OUF) being worked, the labour productivity (LP) and the adjustment factor (PDCF) resulting from the release of standard components to production. In total 71 equations and 35 constants were used.

Vol. 4
1976

3. SIMULATION RESULTS

The obvious basis on which to compare the performance of the system under the two sets of policies is a monetary one but, although such considerations are of major importance, some important factors cannot easily, if at all, be measured in monetary terms. For example fluctuations in the size of the workforce would, because of insecurity of employment, discourage local labour from seeking employment with the firm. Also high levels of overtime for long periods, as well as increasing labour costs, may adversely affect the efficiency of the workforce. Thus a comprehensive assessment of competing policies should include, in addition to monetary measures of performance, a comparison of the time varying behaviour of the system.

For the monetary comparison an average selling price per unit and a simple production cost structure was assumed, the latter including material costs, labour costs, inventory holding costs and hiring and firing costs. Performance was then assessed on the basis of cumulative production costs and cumulative contribution to margin over a period of 3 years, the results showing, for the modified policy case, a decrease in the former of about £100,000 and an increase in the latter of about £350,000. Because the model can only approximate to the real system, too much confidence should not be placed in the absolute values quoted above, but they are large enough to conclude that significant gains can be achieved. In addition a measure of how well the system satisfied demand was obtained by the ratio of the cumulative number of units delivered to the cumulative number of orders expected and on this basis the modified policy case showed an improvement of 5%.

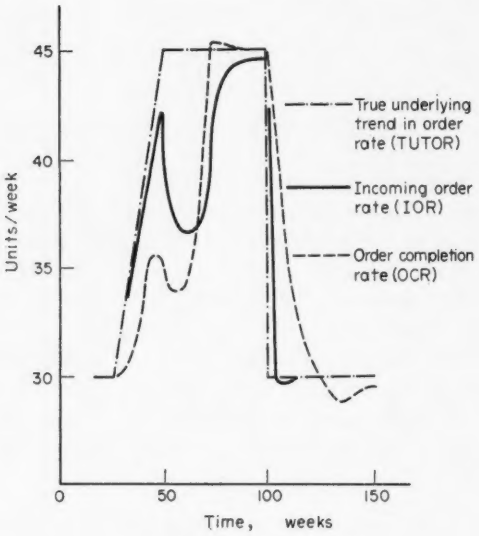


FIG. 2. *Original policies.*

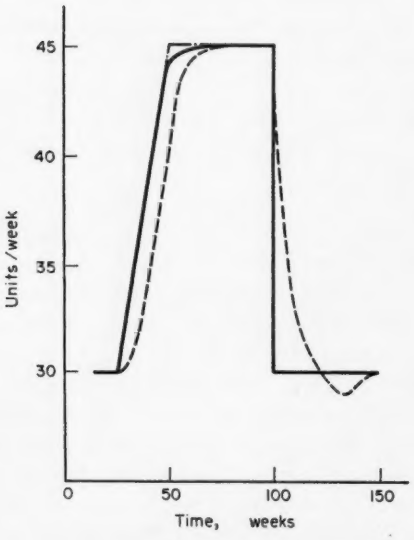


FIG. 3. *Modified policies.*

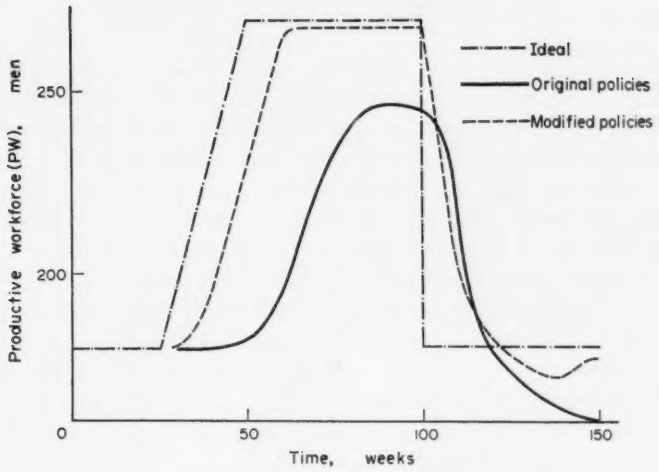


FIG. 4.

Vol. 4
1976

The comparison of the two cases on the basis of behaviour over time is illustrated by the results shown in Figs. 2, 3 and 4. Figures 2 and 3 show the variation of incoming order rate (IOR) and order completion rate (OCR) together with the true underlying trend in order rate (TUTOR), for the original and modified policies respectively. TUTOR is the signal driving the model. On the basis of incoming order rate the preferred behaviour is clearly that associated with the modified policies case since it is scarcely distinguishable from TUTOR. For the original case, on the other hand, significant order losses have occurred as a consequence of awaiting customer complaints to trigger recruitment in response to the increased demand. The behaviour of order completion rate is also superior for the modified system, the divergence from IOR being due to unavoidable delays in the system whereas for the original system, as well as the effect of the increased delays, OCR is additionally distorted, from about week 25 to week 60, by rapidly decreasing productivity consequent upon the excessive build-up of work-in-progress.

Finally, as shown in Fig. 4, the productive workforce in the modified system closely follows the ideal, that is the size of workforce if recruitment in response to changing demand were instantaneous, whereas the build-up for the original case is significantly delayed, never reaching the desired level so that overtime is required to satisfy peak demand. Examination of other system variables showed the modified system behaviour to be superior in all respects and therefore on the basis of both quantitative and behavioural comparisons the modified policies are preferred.

4. CONCLUDING REMARKS

In the introduction to this paper reference was made to the fact that very often management attempt, without external assistance, to assess the probable behaviour over time of their system in response to some expected change and it was claimed that in doing so they were forced to simplify their verbal model to such an extent that their conclusions were unlikely to be valid. Anyone disputing this should attempt to describe the behaviour over time of some selected variables of the system described. They will find themselves quickly overwhelmed by the large number of contributing factors it is necessary to take account of even though this particular case study was limited to the production function of a firm and to a change of policy only with regard to hiring and firing. Even with such a relatively simple system the degree of interaction between the various parts of the system is quite large as can be judged from the information links in the flow diagram and such a degree of interaction makes it impossible to assess realistically the behaviour of the system without the aid of a computer.

As mentioned earlier the system studied was relatively small as may be judged from the fact that only 60 variables were used to describe the system and the computer run, from which the quoted results were obtained, occupied only 1 min 40 sec of central processing time on an IBM 360/65 computer. Obviously much larger systems can be studied without incurring prohibitive computing costs and in general, provided the system is understood well enough to be described verbally, the amount of detail that can be built into the model, and hence its size, is limited only by the capacity of the computing facility available. It is feasible therefore to study systems involving two or more functions of a firm or even the total system of small or medium sized firms.

In conclusion this paper has shown that a detailed verbal description of a system can be translated into a mathematical model which in turn can be analysed over time to enable a choice to be made between competing policies. The realism of the model is dependent upon an accurate description of the system and therefore, if meaningful results are to be obtained, management must be closely involved in the modelling process. Confidence in the method is built up through this involvement and, of course, strengthened by successful application.

5. REFERENCES

1. Forrester Jay W (1961) *Industrial Dynamics.* M.I.T. Press, New York.
2. IBM Corporation (1970) Continuous Simulation Modelling Programme Applications Description Manual (GH20-0240).
3. Lewis CD (1970) Industrial Forecasting Techniques. Machinery Publ. Co., London.
4. Pugh Alexander L (1970) *DYNAMO* II *User's Manual.* M.I.T. Press, New York.

ADDRESS FOR CORRESPONDENCE: *Mr E Stephenson, Department of Business Studies, University of Liverpool, PO Box 147, Liverpool L69 3 BX, UK.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Production Planning in Practice

FL HARRISON

(*Received February 1976*)

The changes and pressures facing the manufacturing and engineering industries today are increasing the importance of effective aggregate manpower and production planning. Several different theoretical optimisation models to tackle this problem have been described in detail in the literature but there have been few applications of them in practice.

The reasons for this are many but include: the difficulty in expressing managements' conflicting and mixed objectives in an objective function; the necessity to oversimplify real life systems to enable these methods to be used; the simplistic approach to manpower planning used in these models; the difficulty in gaining managements' acceptance and finally the fact that what management actually wants is a tool to assist them in planning and decision making.

What is being used by many managements is a case-study deterministic simulation model. Many companies are adopting this type of model for all types of planning and twelve out of twenty-seven companies visited in a research project described in this paper were using this type of model for aggregate manpower and production planning. It is proving to be an effective management tool and is being readily accepted principally because modern specialised financial modelling languages are enabling these models to be built, understood and used by non-specialist managers.

Vol. 4
1976

INTRODUCTION

MANPOWER planning and the balancing of production to meet demand in an optimum manner is difficult, particularly when sales have seasonal variations. In addition, when demand is subject to major change, efficient aggregate manpower and production planning can be vital to profitability, good industrial relations and survival. This is particularly so under recessionary conditions as in 1974–76, when many managements are evaluating alternative combinations of short time working, partial closures and redundancy.

One method of tackling this problem is by using a computer based aggregate production planning model. This takes as inputs the sales forecasts and management's policies and then attempts to optimise the decisions necessary to balance manpower and production to achieve management's objectives. In aggregate production planning, decisions must be made for each period with regard to size of labour force, amount of overtime or short time working, level of stock

holding and amount of sub-contracting. All of these have costs and non-quantitative implications which interact in a complex manner so that the optimum set of decisions is often difficult to determine.

Several different theoretical aggregate production planning models have been written about in detail, but there have been few applications of them in practice. This paper discusses some of the reasons for this and also one method of aggregate manpower and production planning which was observed to be in widespread use during research on a Social Science Research Council project on the use of computer models.

LITERATURE MODELS

Five types of aggregate production planning models have received considerable attention, viz:

(1) Linear programming;
(2) Linear decision rule;
(3) Parametric production planning;
(4) Management coefficients model;
(5) Search decision rule.

The first two methods depend on mathematical programming and the third on a heuristic approach with two linear feedback rules for labour and production rates and a computer search routine. The management-coefficients model is based on decision rules and a statistical analysis of past management decisions. The Search Decision Rule involves a simulation model of the system and a computer based optimum search procedure to optimise an objective function.

There have been few applications of these methods in industry and most reported examples have involved fairly simple production systems.

These methods have the following drawbacks:

1. They generally require an objective function to be established and it is normally not easy to express management's desires and objectives in a straightforward mathematical objective function. Management's objectives are often multiple, conflicting and a mixture of quantitative and non-quantitative factors. For example, they may want to minimise overall operating costs and working capital tied up in inventory; reduce work-force fluctuations and work some policy overtime to encourage good labour relations. All these are essentially conflicting objectives.

2. These models tend to tackle homogeneous labour forces and products. Most production situations are much more complex than the simple examples illustrating use of these models, with multi-product, multi-

departmental structures and varying product mix requirements each period. This leads to varying departmental manpower needs with normally different trades requirements. The problem is complicated further when the model has to handle inter-departmental stocks and the fact that products required in time period t require labour, materials, sub-contracted components and equipment from different departments in time periods $t - i$, $i = 1, 2, 3 - - n$. In addition, most sub-contracting is for intermediate components on a make or buy decision rather than for finished goods.

3. They tend to vary the labour force in a simplistic hire and fire basis whereas most managers are reluctant to operate in this manner because of its adverse influence on labour relations. Complex work rules are required to express management's practical policies in this area which must take into account such factors as: using natural wastage to reduce manpower when required, time to train new workers, financial effects of reduced working weeks on the allocation of overheads, the necessity to negotiate and justify to unions and workpeople any redundancies or reduced work-week, overtime policy and policies as to hiring and firing. For example one firm stated that it only hired new workers if they could ensure continuity of working and only fired men if the downturn in demand appeared to be prolonged indefinitely after they had stocked as much as they reasonably could and operated a reduced working week for some time.

4. These models tend to ignore cash flow problems, that is working capital tied up in inventory. In times of cash flow shortage and high interest rates, management will often accept increased operating costs, that is overtime or sub-contracting, in order to reduce cash flow peaks.

5. Most aggregate planning models in the literature are purely production models while in practice they need to be integrated with the marketing and financial operations to be useful management tools.

The Search Decision Rule model is the most flexible of the above methods and can handle some of these factors, but unfortunately it, along with the others, suffers from great difficulty in gaining management acceptance. The mathematics and computer programming required for these models necessitates specialists to build them and most managers do not understand the mathematics involved. This leads to their placing little trust in these models and their reluctance to accept the output. Because generally the real-life production situations are too complex to be easily handled by the techniques, they normally have to be simplified in the model and this sometimes simplifies out the reality. In addition, without gaining management's understanding and co-operation, the quantitative specialist may find difficulty in correctly analysing the production situation.

Vol. 4
1976

Thus not only will there be a probability of the model not being realistic, but almost certainly without gaining management's understanding and acceptance the model will not be used in practice.

Another reason for this lack of acceptance is that the type of model which is most useful to the manager in planning is not necessarily an optimisation model. Planning is a decision making process and not a straightforward optimisation process. It involves the search for alternative courses of action, the tightening and slackening of constraints, judgemental trade-off between objectives and what is desirable and feasible. The generation of alternatives is often carried out sequentially involving a balancing of quantitative and qualitative factors; a process of building and evolving as the manager moves from one compromise to another. Optimisation is not excluded but is more likely to involve a limited search for the optimum value of a sub-set of the variables subject to specific constraints.

Instead of an optimisation model many managers are adopting a straightforward deterministic simulation model that assists them in their planning and decision making but which does not necessarily take the decisions for them. This type of model simply carries out quickly the large amount of calculations required to evaluate the consequences of any set of decisions or plan. As such it can be critically called merely a 'number cruncher'.

It is, but the man–model combination also greatly enhances the thinking ability of the manager in planning and decision making. In simpler problems, a manager may, during his analysis, carry out some calculations to evaluate one line of thought, discard it and try another. Then based on this analysis he may conceive of another line of thought leading to an eventual solution which was not amongst his original alternatives.

With a deterministic model structured around his production system the manager can carry out the same mode of thinking and evaluation but on much larger and more complex situations. One manager interviewed commented that the man–model combination can almost add another dimension to the thinking ability of managers. Correctly used it can be a much more flexible and practical tool than an optimisation model.

THE CASE STUDY DETERMINISTIC MODEL

In the last few years there has been an explosion in the number of firms using computer models in their planning and decision making. In a survey carried out in 1974, the computer manufacturers and bureaux claimed that 1000 companies in the UK were using their modelling systems.

Though published examples have emphasised corporate simulation models, 18 out of 27 of the companies observed in this research project used case study deterministic computer models for operational planning and decision making.

Firm	Application
Furniture and flooring company	Manpower and facility planning: including the closure of one factory due to reduced demand.
Furniture company	Manpower and facility planning: including plant rationalisation.
Television rental company	Manpower planning in television maintenance: seasonal maintenance demand and capital requirements, i.e. a car for every man.
Passenger transport authority	Manpower planning and control, maintenance budgeting, facility and training planning.
Management consultant company	Manpower planning.
Food processor company	Manpower, material and production planning: model used to allocate short supplies during three-day week working.
Plastics company 'A'	Manpower and extrusion machine planning: highly seasonal demand combined with major stocking problems, machine allocation and occupation troubles.
Plastics company 'B'	Manpower and facility planning: sales planning and control, allocation of orders to plant, purchasing planning.
Chemical company	Multi-plant, intermediate components and multi-product production planning.
Agricultural machinery company	Manpower and production planning for short-run production line.
Industrial equipment company 'A'	Manpower and production planning for multi-product, multi-trade, multi-department factory.
Industrial equipment company 'B'	Manpower and production planning: as above plus multi-project planning.

FIG. 1. *Firms using aggregate production planning models.*

Twelve of these companies had built and were using aggregate manpower and production planning models (see Fig. 1).

The case study deterministic model can be built to represent the production system as it actually exists with the only simplification being the level of aggregation used. It is structured around the production facilities, identifying the main elements and how they relate to each other and the variable factors involved. These relationships are expressed as algebraic equations. They represent in mathematical form, the constraints and how the production system operates, and they can be used to establish how the system behaves as these key factors

Vol. 4
1976

are varied. The variables in these equations are the variables with which the manager is accustomed to work namely, sales, stock levels, planned production levels, outputs and cost control estimates, etc.

These models show some similarity to the Search Decision Rule model but are used in an entirely different way. This deterministic type of model is normally much more realistic than the literature models and can handle more complex situations. There is no need for an objective function in that this simulation model can present alternative solutions with various compromises of mixed objectives for the manager's final judgemental decision.

Basically, the model is a computer programme which calculates out the production, manpower and financial results of alternative production plans to meet sales demand. It does this by simulating what would happen with various departmental machine capacities, tradesmen and other labour levels, product mixes, stock levels, sub-contracting policies, overtime, short term or shift working. It is not concerned with allocating individual jobs to machines as in production control systems but with aggregated load hours on machine centres or departments for each period, balancing demand against capacity.

It can be used, for example, with seasonal demand to calculate out the results that would be obtained with the existing manpower levels, and various overtime and sub-contracting strategies. The manpower levels could then be varied on the basis of expressed management policies to examine what would be the results with these various strategies.

In recessionary conditions, when the future level of demand is extremely uncertain, it can be used to evaluate the consequences of various manpower strategies with different forecasts, e.g. optimistic, pessimistic and most likely. If management wished to operate a policy of no forced redundancy and use natural wastage to reduce manpower, the costs and practicality of this could be simulated.

If production still outstripped demand and the resultant stock levels could not be carried or financed, the model could be used to examine what level of short time working would balance them. It may then still be necessary to examine various plant closures and redundancy schemes before production and demand can be balanced.

These can all be simulated with such a model and the results used in conjunction with non-quantifiable factors to assist management in planning and decision making. The cost of various management policies can be balanced against other factors. The fact that such evaluations have been carried out and can be demonstrated can be of considerable assistance in negotiations with the trade unions and Government departments who are increasingly becoming involved in such situations.

Similarly, in times of rapid inflation working capital requirements can be greatly increased. This type of model can be used to simulate various production plans with, for example, reduced stock holdings and increased overtime and

sub-contracting to examine alternative ways of reducing working capital requirements.

This type of deterministic model is able to evaluate quickly many alternative sets of decisions and situations by the use of data files for each set of circumstances. The input data for each alternative scenario or case is stored in a data file unique to that combination of forecasts and decisions. The model can then be run to evaluate each alternative case simply by specifying which data file is to be used as input.

This enables planning to be more of a continuous process and to be more analytically based than if the evaluations were carried out manually. Without a model the calculations involved in each case would often take a significant time. The manager would probably start with the calculations involved in a few initial alternatives; these would suggest new alternatives which would involve further delay while these calculations were carried out. By the time the results were produced, management would have lost its train of thought and perhaps forgotten what the assumptions for that case were.

In practice, without a model, this type of planning is rarely comprehensive or analytically based. A few alternative plans would be evaluated and if one balanced production with sales, it would be adopted. This allows the planning process to be based on systematic analysis without unacceptable delay.

Vol. 4
1976

In addition, the deterministic model can be used for analysis and control of progress. Rarely, if ever, does everything go according to plan. This type of model can also be used to carry out planned vs. actual comparisons and various analysis. It can then be used to evaluate the consequences of any changes and to simulate alternative revised plans or decisions.

In several of the firms visited the aggregate planning models were completely independent models but in others they were integrated into the overall operations planning models of the company. These operations planning models, although nominally financial models, in fact contain marketing, production and financial sub models with only money as the common denominator. The output from the marketing models feeds into a data file which is the input file for the production planning model, which, in turn, feeds into the data file for the financial sub model. This integration avoids the dangers of functional sub-optimisation.

A major factor in the acceptance of these models is that management is learning to build and use them by itself with little or no specialist assistance. These models are relatively straightforward to build as they are simply models of the actual production system with no artificial or mathematical complications and the inputs and outputs are in a form familiar to management.

This is particularly so when a specialised financial modelling language is used to build them. These are now widely available from many computer manufacturers and bureaux. Though designed to permit non-data processing and non-quantitative accountants and managers to build and use financial models,

their structure is such that they are ideally suited to building aggregate production planning models. This is primarily due to their automatic handling of multi-period situations, near normal English language instructions and easy consolidation of sub models. There is no doubt that a major factor contributing to the increased use of computer models is the fact that the non-computer specialist can quickly build, understand and use computer models with these modelling languages.

CONCLUSION

The theoretical models developed in the literature for aggregate production planning have not been adopted to any significant extent by industry. Instead, industry is adopting the case study deterministic simulation model to carry out the functions of aggregate manpower and production planning. The changes and pressures facing the manufacturing and engineering industries today are increasing the importance of this mode of planning and the case study deterministic model is proving to be an effective tool to cope with the practical problems involved.

This type of model is capable of handling complex systems, does not require a mathematical objective function, can incorporate management's desires and policies and is capable of being built, understood and used by the non-specialist manager. Its principal difference from the theoretical models described in the literature is that it is used as a tool to assist management in planning and decision making, rather than as an optimising model, although it can be used for a limited search for a constrained optimum.

ADDRESS FOR CORRESPONDENCE: *Mr FL Harrison, Department of Management Studies, University of Glasgow, Adam Smith Building, Glasgow, G12 8RT, UK.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

An Application of Multiple Objective Linear Programming to Media Selection

RALPH E STEUER

RICHARD L OLIVER

(*Received August 1975; in revised form March 1976*)

A multiple objective linear programming (MOLP) method utilizing interval criterion weights is applied to the problem of media selection. Using this technique, it is possible to analyze a problem more explicitly in terms of the several objectives inherent in many media selection situations. In order to illustrate the interval criterion weights approach, a multiple objective hierarchical media selection model is presented and its computer results discussed. In addition to being able to deal more directly with different decision criteria, a distinguishing feature of the mathematical analysis applied here is that its output enables the media-planner to be presented with a small cluster of candidate media schedules (rather than just one). Then, from this list, the media-planner should be in a position to qualitatively make a final choice as a close approximation to his optimal solution.

Vol. 4
1976

INTRODUCTION

THE PURPOSE of this paper is to discuss the application of a newly developed technique, Multiple Objective Linear Programming (MOLP) with interval criterion weights [6], to a multiple criterion media selection problem. This technique actually functions as a generalization of regular linear programming in that it can process problems in which more than one objective function is specified. Consequently, users are no longer necessarily limited to a single composite objective function when more than one different measure of effectiveness would be more appropriate as is often the case. Another feature of this solution method is a provision whereby the media planner may associate an interval weight with each different objective. This capability can be applied to enable the procedure to reflect two types of uncertainties that may be present in the problem formulation. The first pertains to the relative importance placed upon the different criteria when evaluating candidate media schedules, while the second pertains to errors that may exist in the estimated values used as

coefficients of the objective functions. As a result of these capabilities of the multiple objective procedure, particular types of media selection models may be more amenable to analysis than has previously been the case.

The history of multiple criteria advertising effectiveness models dates from 1925 when the AIDA model (attention, interest, desire, action) was originally presented. Later modifications resulted in a five stage model where conviction or memory was added as a fourth stage between desire and action. Subsequent to these initial variations of AIDA, a number of multiple stage or 'hierarchy of effects' advertising models appeared around 1960 including those in [1, 3, 7]. Since that time, many new or refined hierarchial models have been proposed including the recent Howard and Sheth [2] model where the stages are conceptualized as intervening variables.

A number of underlying assumptions are common to the various models cited above, the most basic being that potential consumers can reasonably be expected to move through stages in response to advertising appeals. Moreover, it is assumed that this progression is unidirectional and that higher level stages denote greater purchase probabilities. From the perspective of defining distinct advertising goals, however, the most important assumption is that an advertising strategy can influence movement to each hierarchical stage and that particular media and/or message content vary in effectiveness at different stages. The remainder of this paper will draw on the widely cited 'hierarchy of effects' model proposed by Lavidge and Steiner [3] as an illustration of a multiple objective linear programming approach to media selection.

Vol.
197

THE MULTIPLE OBJECTIVE ADVERTISING EFFECTIVENESS MODEL

Lavidge and Steiner posited that potential consumers move through seven attitudinal stages in response to advertising appeals. Respectively, these steps are unawareness of the product or service followed by awareness, knowledge, liking, preference, conviction, and purchase. Because some consumers will reside at every level at one point in time (except perhaps for an innovation), an upward movement from each stage can be viewed as a separate and distinct advertising objective. Thus, a media mix will be more desirable to the extent that it can influence movement to all hierarchical levels to varying degrees. In addition, a multiple objective approach to promotional effectiveness will permit the media planner to explicitly account for the differential effects of various advertising and promotion appeals. Lavidge and Steiner originally proposed that different vehicles be used for different advertising goals and made predictions as to which of the six stages following unawareness would be most affected by a variety of promotions. Their example is shown in Table 1.

TABLE 1. OBJECTIVE-SPECIFIC PROMOTIONAL STRATEGIES*

Specific objective	Promotional strategy
Purchase	Point-of-purchase displays, retail store ads, deals, price appeals
Conviction	Testimonials
Preference	Competitive ads, argumentative copy
Liking	'Image' ads, status or glamour appeals
Knowledge	Announcements, descriptive copy
Awareness	Classified ads, slogans, jingles

* Adapted from [3, p. 61].

A MOLP model for such an hierarchical model might be as in (1)

$$\max \left\{ \sum_j w_{ij} x_j = z_i \right\} \text{ for } i=1, \ldots, K, \tag{1}$$

$$\text{s.t. } \sum c_j x_j \leqq B,$$

$$0 \leqq x_j \leqq b_j \text{ for all } j.$$

However, since such unbridled formulations typically generate unmanageably large numbers of efficient (i.e. Pareto optimal or nondominated) solutions (see [6]), a modification involving interval criterion weights being associated with the different objectives

$$\lambda_i \epsilon (l_i, \mu_i) \text{ for all } i, \tag{2}$$

$$\sum_{i=1}^{k} \lambda_i = 1$$

where:

- x_j is the number of units of the jth medium;
- w_{ij} is the unit effectiveness coefficient of promotional medium j for objective i;
- c_j is the unit cost of medium j;
- b_j is the maximum allowable number of units of medium choice j in any feasible media schedule;
- B is the total advertising budget.

will be incorporated into the problem in order to reduce the amount of information generated to workable proportions. Thus the hierarchical advertising

Vol. 4
1976

model to be treated in this paper will now consist of the MOLP formulation (1) with interval criterion weights (2).

In formulation (1) it has been assumed that through prior behavioral analysis sufficient information has been acquired in order to make reasonable estimates of the objective-specific effectiveness coefficients w_{ij}. Also, it has been assumed that the criterion weight intervals (2) have been appropriately selected to reflect (a) the relative emphasis to be placed on respective movements to higher cognitive levels and (b) any possibilities for error in the w_{ij} estimates. In other words, the purpose of the intervals (l_i,μ_i) is to generate all extreme points from that portion of the efficient surface of the feasible region in which the data of the problem suggest the optimal point may lie.

When using the example of Lavidge and Steiner, the z_i in (1) would be awareness, knowledge, liking, etc. while the $x_{j,j}$ small, would represent those media choices most effective for the early cognitive stages (e.g. 'jingle' campaigns) whilst $x_{j,j}$ large, would be reserved for action oriented strategies (e.g. point-of-purchase displays). Note that, while particular media are hypothesized to affect certain criteria more than others, it is expected that other hierarchical stages will experience lesser 'residual' effects. These effects will be more pronounced for stages immediately preceding or following the one for which a promotional effort was designed and weaker for more remote levels of cognitive development.

The procedure for solving for all of the extreme point solutions of the MOLP problem (1) with interval criterion weights (2) is drawn from the methodology presented in [6]. Briefly, this involves solving the family of weighted-sums problems (3)

$$\{\max [\lambda^T W x : x \in S] \mid \lambda \in \Lambda^1\} \tag{3}$$

where:

W is the criterion matrix formed by the w_{ij} coefficients;
S is the feasible region $\{x \mid \sum_j c_j x_j \leqq B, 0 \leqq x_j \leqq b_j\}$;
Λ^1 is the set $\{\lambda \mid \lambda_i \in (l_i, \mu_i), \Sigma^k_{i=1} \lambda_i = 1\}$

for all maximizing extreme points. Although not directly available, these points can be obtained, in turn, by solving an equivalent vector-maximum problem, 'max' $\{Dx|x \in S\}$ where D is a new criterion matrix, for all efficient extreme points (see [6] for complete details).

HYPOTHETICAL PROMOTION SELECTION EXAMPLE

The numerical example of this section has been constructed to illustrate an application of MOLP with interval criterion weights in promotion selection. In

practice, a user might wish to add refinements including nonlinearities (to be handled in a piecewise linear fashion), time and space considerations, overlapping audiences, etc. but for purposes of exposition the simple model presented here should suffice.

The problem involves four of the six criteria discussed in [3] and nine promotional media variables selected for their objective-specific effects. In each case, the objective value departs from the traditional use of exposure units and is expressed in terms of the number of consumers induced to move from a lower hierarchical stage to a higher level of cognitive development. The various criteria are discussed below.

Awareness. This objective is the most direct analog of the traditional exposure criterion in that it represents the number of new consumers who have become aware of the manufacturer's product or brand as a result of a promotional effort. Network radio and TV jingle/slogan campaigns were selected as the media and messages most likely to influence the awareness objective although residual effects on other criteria would be expected.

Knowledge. One consumer's acquisition of product or brand related knowledge would represent a one unit increment in the knowledge criterion. Ordinary non-competitive descriptive copy promoting the merits of the advertised product would best achieve this objective although, as in the case of the awareness criterion, other objectives would experience lesser residual effects. Both network TV and a representative national weekly magazine were selected to promote the descriptive copy.

Vol. 4
1976

Preference. The preference objective is unit incremented when one consumer acquires a preference for the product or brand of interest. Competitive copy (favorably) comparing the advertiser's brand to competitive brands is considered as having its greatest effect on this criterion. The same TV and weekly magazine media used for the knowledge criterion were used here as well.

Purchase. The purchase objective is defined simply as units sold (assuming one per customer). Retail store copy in the local media was chosen to best influence this criterion. The selected media included spot radio, spot TV, and local newspaper promotions.

As was noted, the objective–specific media coefficients (i.e. the w_{ij}) required for a MOLP analysis represent the number of potential consumers moved to each hierarchical stage as a result of one unit of promotion x_j. Moreover, because each promotional strategy is assumed more effective at certain cognitive levels, the media coefficients for the same promotion will vary over the criteria.

Admittedly, determination of actual values of the objective–specific promotion-effectiveness coefficients might involve a substantial data collection effort. However, the demands for precision normally required in the calculation of media coefficients need not be as great with the current solution approach because any uncertainties in the objective-function coefficient appraisals can

TABLE 2. PROBLEM COEFFICIENTS FOR MEDIA SELECTION EXAMPLE: TOTAL BUDGET = $1,000,000

	Slogan/jingle campaign		Descriptive copy		Competitive copy		Retail store copy		
	Network radio	Network TV	National magazine	Network TV	National magazine	Network TV	Spot radio	Spot TV	Top 100 newspapers
z_1: Awareness	15,000	50,000	10,000	10,000	10,000	10,000	5000	5000	5000
z_2: Knowledge	1000	5000	20,000	35,000	15,000	25,000	1000	1000	1000
z_3: Preference	0	0	15,000	25,000	20,000	35 000	1000	1000	1000
z_4: Purchase	0	0	1000	1500	1500	2000	5000	10,000	20,000
Variable	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Constraint	26	26	12	12	12	12	6	6	6
Unit cost	$2300	$25,250	$39,190	$50,500	$38,190	$50,500	$3250	$30,900	$121,000

be integrated into the problem by widening the interval criterion weight widths. This would typically result in an increase in the size of the neighborhood of efficient extreme point media schedules generated thus increasing the likelihood that the set of points enumerated will contain the media-planner's 'best' promotion schedule. From this cluster of schedules, the media-planner should qualitatively be able to evaluate the list and identify the most preferable solution.

Due to the unavailability of actual data, the coefficients used in the present numerical example were selected by the authors as being representative. They reflect the assumptions underlying the choice of media in that certain promotional strategies are believed to differentially influence the criteria. The numerical example analyzed in this paper is described in Table 2. Unit cost data were obtained from [4, pp. 713–723]; the budget was set at $1,000,000; and the media constraints (the maximum allowable number of annual medium insertions) were chosen as similar to those used in prior studies.

To show how the problem in Table 2 might be approached for two pertinent competitive strategies, illustrative interval ranges for the criterion weights are given in Table 3. The first weighting scheme was constructed to suggest the

TABLE 3. CRITERION WEIGHT INTERVALS USED IN TWO PROMOTIONAL STRATEGIES

Strategy	Objective z_1	z_2	z_3	z_4
1. New unique product	(0·3, 0·7)	(0·2, 0·5)	(0·0, 0·1)	(0·1, 0·2)
2. Mature competitive product	(0·0, 0·1)	(0·0, 0·1)	(0·4, 0·7)	(0·4, 0·7)

criterion emphasis an advertiser might select for a relatively new product with a market monopoly. New awareness would be his primary goal although product knowledge would also be important. Due to the unique nature of the product, emphasis on purchase promotions would be relatively light while emphasis on preference would be unnecessary. In contrast, the second scheme reflects a mature market environment where brand awareness and knowledge are assumed. However, because the market for the advertiser's product would ordinarily be competitive, the preference and purchase criteria would acquire large weights.

Candidate solutions

The example in Table 2 was solved for the two interval weighting schemes using computer algorithm ADBASE [5]. The two sets of efficient extreme point media schedules obtained are presented in Table 4 where it can be seen that strategy 1 generated two candidate solutions while strategy 2 generated four.

Recall that strategy 1 concerned a new product with a market monopoly where the criterion weight intervals were selected to emphasize awareness and

TABLE 4. SETS OF EFFICIENT EXTREME POINT MEDIA SCHEDULES FOR STRATEGIES 1 AND 2

Item	Extreme point media schedules					
	Strategy 1		Strategy 2			
	1	2	1	2	3	4
z_1: Awareness	1,772,316	1,789,180	622,403	602,314	248,062	224,158
z_2: Knowledge	345,108	300,360	455,605	550,108	453,093	565,554
z_3: Preference	136,792	109,770	590,807	581,792	622,124	611,396
z_4: Purchase	37,847	36,918	66,360	63,347	68,709	65,123
x_1	26·000	26·000	26·000	26·000	—	—
x_2	26·000	26·000	—	—	—	—
x_3	—	6·918	—	—	—	—
x_4	5·231	—	—	6·231	—	7·415
x_5	—	—	8·240	—	9·806	—
x_6	—	—	12·000	12·000	12·000	12·000
x_7	6·000	6·000	6·000	6·000	6·000	6·000
x_8	—	—	—	—	—	—
x_9	—	—	—	—	—	—

knowledge over the other criteria. To satisfy these objectives, both candidate solutions for this situation fully utilize media 1, 2, and 7. The solutions differ in their use of medium 3 (the national magazine) as opposed to 4 (network TV); the first schedule used TV to the exclusion of the print media while the second selected print over TV. To decide between these media choices, the decision-maker would need to compare the criterion values generated by each solution. An inspection of Table 4 shows that the magazine solutions emphasize the knowledge and preference criteria while the TV solutions emphasize awareness. In effect, the media-planner would be 'trading off' roughly 17,000 new awareness units for 45,000 new knowledge and 25,000 new preference units. Although this decision may appear difficult to some and easy to others, a media-planner would conceivably be able to resolve the situation by making final reflections upon his present media purposes.

Criterion weights for strategy 2 were selected as representative of those that might be employed for a mature product. The preference and purchase objectives were assigned large criterion weights while very small weights were applied to the first two objectives. Table 4 shows that the set generated by these criterion weight intervals contains four efficient extreme point media schedules. All were similar in that they fully utilized media 6 and 7. They differed in that schedules

Vol. 4
1976

1 and 3 also utilized medium 5, solutions 2 and 4 selected medium 4, while solutions 1 and 2 fully utilized medium 1.

The differences between the four choices are, of course, reflected in the criterion values. All solutions yield approximately the same level of new purchases while schedules 1 and 2 result in considerably more new awareness units than schedules 3 and 4 although the latter yield greater preference levels. Finally, solutions 2 and 4 result in a higher level of knowledge units. The media-planner would wish to take these differences into consideration when making his final selection from among the four candidates.

CONCLUDING REMARKS

The advantage of using the interval criterion weights technique in place of a standard linear programming approach to media selection is that the former typically generates a cluster of candidate solutions about the single recommendation that would normally result with the latter. Rather than being presented with just a single solution, the decision-maker is thus able to make his final selection in the light of all neighboring extreme points corresponding to his uncertainties about the actual criterion weights and objective function coefficients as specified via the intervals.

Vol.
197

REFERENCES

1. COLLEY RH (1961) *Defining Advertising Goals for Measured Advertising Results.* Association of National Advertisers, New York.
2. HOWARD JA and SHETH JN (1969) *The Theory of Buyer Behavior.* Wiley, New York.
3. LAVIDGE RJ and STEINER GA (1961) A model for predictive measurements of advertising effectiveness. *J. Marketing* **25**(4), 59–62.
4. MANDELL MI (1974) *Advertising*, 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey.
5. STEUER RE (1975) ADBASE: an adjacent efficient basis algorithm for vector-maximum and interval weighted-sums linear programming problems. *J. Marketing Res.* **12**, 454–455.
6. STEUER RE (1976) Multiple objective linear programming with interval criterion weights. *Mgmt. Sci.* In Press.
7. *Toward Better Media Comparisons* (1961). Advertising Research Foundation, New York.

ADDRESS FOR CORRESPONDENCE: *RE Steuer, Department of Business Administration, College of Business and Economics, University of Kentucky, Lexington, Kentucky 40506, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Job Shop Scheduling—A Case Study[1]

PAUL H RANDOLPH

(*Received October 1975; in revised form March 1976*)

A missile range is essentially a large-scale job-shop, involving prodigious amounts of test equipment and formidable problems of coordination. Because of its characteristics, a missile range can be considered a one-machine, *N*-job situation. For this problem the major scheduling methods were examined and tested, but all were discarded as unfeasible except one which used a Monte Carlo scheduling procedure adjoined with statistical stopping rules. An algorithm based on these ideas was constructed, and it has proven to be flexible and workable, providing predictably near optimum schedules for the missile range within a probabilistic and statistical framework. Even though there is not total implementation yet, considerable benefits to the missile range have already been experienced. For example, in order to automate the scheduling process, it was necessary to codify the goals of the range, something that had never been formalized before. This involved an unprecedented scrutiny and precision of definition of these goals, and from this study a reasonable numerical optimization criterion was constructed. Also, in order to use any automated scheduling algorithm, data files had to be stored on tape for ready access, which in turn has improved other operations on the missile range that were dependent on these data. Furthermore, the scheduling algorithm is providing conflict-free schedules in a few minutes of computer time.

Vol. 4
1976

INTRODUCTION

SCHEDULING problems of various types have been in existence for many decades. However, there is one scheduling problem that has literally been introduced by the space age: the scheduling of a missile range.

A missile range is a highly complex and extensive operation, one that gives rise to unique scheduling problems. For example, the sheer size of a missile range presents extraordinary problems of coordination. Also, on a missile range all operations of a job are performed simultaneously rather than sequentially; in this sense it is analogous to a one-machine shop.

Scheduling a missile range is a daily process in which twenty to forty jobs are to be scheduled. Since manual methods of constructing schedules are difficult

[1] This paper received the 1974 Award of Military Applications Section of Operations Research Society of America.

and time consuming, a computerized method should eliminate most of the difficulties and result in considerable savings. Although this paper is a description of some of the problems associated with the development and installation of a workable approach to that problem at a particular missile range, it should be possible to adapt the procedure to all missile ranges. In fact, the general concept is applicable to almost all job-shop scheduling problems.

MISSILE RANGE SCHEDULING PROBLEM

Since a missile range is a large-scale job-shop, manual scheduling methods similar to those in industry have been used at each missile range. Essentially, this generally is a committee effort in which representatives of various branches of the service meet with personnel of the scheduling office to develop a range schedule for a particular day. Normally, this development process is extended over a period of several days, and begins about a week ahead of time. On the first day of schedule construction the various branches present their job requirements including preferred time-of-day for each job, if applicable. For each job, there exists a 'route sheet', a list of various resources or 'machines' to be used for the job. These resources are used not only during the time the mission or job is being run, but, in addition, they must be readied before each job and, after the job, converted or closed down. For example, fixed cameras mounted close to a launch pad must be ready to run at least half an hour prior to the launch time—well before a missile is readied for flight—and then the immediate area must be evacuated for safety reasons. Conversely, such items as, for example, telemetry and radar devices may be required for some time after the launch, not only to record the complete flight of the missile but also partially to process the data. These requirements are illustrated in Table 1, where the *T*-time is the scheduled launch time. As can be seen, the various resources have differing beginning and end times.

There may also be certain other restrictions on the resources. For example, it may be necessary to restrict the mission to a particular time interval in the day—called a 'window'—as may be the case for jobs using cameras dependent on the position of the sun. Each machine has a certain 'set-up-time', that is, the time interval required to change from one job to another. It is estimated that a radar, for example, requires about half an hour for its housekeeping chores: remove data tapes, insert new tapes, ready plotting boards, calibrate, etc. Also, it may be necessary to transport portable equipment from one site to another, sometimes a distance of several miles. These 'turn-around' or 'set-up' times must be considered in any scheduling job at a missile range.

In manual methods of scheduling the information for each job is summarized in coded form on plastic strips which are about 4 ft long and 3 in. wide. These strips are then attached to a large board in the approximate order, from left to

TABLE 1. RESOURCE REQUIREMENTS FOR SCHEDULED MISSION

Index	O D Number	Short title	*T*-Time
AK	81321E	Bonneville	910

Prime resources

Resource number	Beginning time	End time	Number used
E30	900	915	1
F114	900	915	1
T128	905	915	1
J26	905	925	1
D55	910	920	1
F392	910	920	1
F393	910	925	1
G28	905	920	1
G106	910	925	1
G107	910	925	1
J1	905	925	1
J10	905	925	1
R125	905	925	1
J62	905	925	1
T865	905	920	1
T866	905	920	1
T290	905	920	1
D59	905	925	1
D63	905	925	1
D68	905	925	1
D68	905	925	1
D69	905	925	1
D70	905	925	1
T301	910	925	1
TTAS	910	925	1
F401	910	925	1
R112	905	920	1
R123	905	920	1
F412	900	925	1
R122	905	925	1

Non prime resources

Resource number	Beginning time	End time	Number used
G235	905	925	1
T301	910	925	2

right, according to the time the jobs would be run if the requested time-of-day for the job were observed. As might be expected, this first approximation to the schedule usually results in gross conflicts among resources, often with excessive overtime being scheduled. From this first approximation, the plastic strips are then interchanged and new times are tried. Each time a new permutation is presented, extensive search is needed to determine if the resulting schedule is feasible; that is, if no conflicts exist. This trial and error procedure may continue for several days, until a schedule is obtained that, hopefully, is both

Vol. 4
1976

feasible and also exhibits some degree of optimality; that is, minimizing overtime and, at the same time, scheduling as many of the requested jobs for the day as is possible. The stripboard is essentially the Gantt chart of missile range scheduling.

The manual strip manipulation process of scheduling points out, however, that a missile range is a one-machine, N-job structure, and that the concept of a permutation of the day's jobs for the first step in the scheduling process, basic to the one-machine N-job problem, can be utilized in a computerized scheduling algorithm.

COMPUTERIZED METHODS OF JOB-SHOP SCHEDULING

When the missile range scheduling problem was examined there seemed to be four possible approaches towards a solution: mathematical programming, implicit enumeration, simulation, and Monte Carlo. Each of these were examined and all but the last were rejected as unworkable.

The use of mathematical programming for the job-shop scheduling problem was probably first suggested by Wagner [13] in 1959. However, this method, as well as the others that followed, were unwieldy and very slow even for small sized problems. Thus, after brief considerations the approach was abandoned.

Many individuals have felt that implicit enumerations would be the desirable way to solve scheduling problems [3]. The argument is that implicit enumeration would require the examination of only a fraction of the total number of possible schedules. We thought we might test this method by taking a small problem and first finding all possible permutations. We discovered that even a nominal 23-job problem would require 23! schedules for a complete enumeration. If each schedule required only 1 μsec to develop on the computer, it would take more than the astronomical age of the universe to obtain all 23! possibilities, assuming the computer ran 24 hr a day.[2] Since any measurable percentage of this would be far too long, we abandoned implicit enumeration.

Simulation has been suggested by Fenema [5], in which the efforts of human schedulers are simulated on a computer. Reports from the places in which simulation procedures have been installed are not encouraging. The computer logic becomes very cumbersome and lengthy, and the programs take excessive amounts of computer time. Furthermore, the computer lacks the 'gut' feeling, the intuitive recognition, of optimality that the simulated humans would have. Simulation was also crossed off the list.

This left Monte Carlo methods [1, 6, 8] in which schedules are generated at

[2] The author is indebted to Dr. Robert Machol of Northwestern University for supplying this dismaying piece of information.

random and after 'enough' schedules have been obtained, the best of them is retained. Chances of achieving a near optimal solution are good if enough schedules are generated. The primary advantages of a Monte Carlo approach is that a solution for a large complex problem can be obtained in a reasonable length of time. Although this may also be true of other scheduling techniques such as a truncated implicit enumeration [2], Monte Carlo becomes especially attractive when combined with the modern statistical theory of stopping rules. This combination provides a very efficient and flexible scheduling algorithm; it not only indicates the point at which the Monte Carlo sampling process is to be stopped, but, almost as important, yields a statistical measure of the quality of the schedule that is used.

THE MISSILE RANGE SCHEDULING ALGORITHM

For this one-machine, N-job scheduling problem, a Monte Carlo algorithm was constructed, based on the following simple sequential rule: a random listing of the N-jobs is generated and the jobs are scheduled in the order of the permutation as early in the day as possible (consistent with a time-window if this is a requirement), but without conflicts or overlaps in the use of instrumentation. Because of the limitations in the working day and because of the great potential for conflicts in the resources, it is likely that some jobs will not be scheduled. Another random permutation of the N-jobs is generated and the scheduling procedure is repeated. In general, a different set of jobs is scheduled. The repetitive process of generating random permutations and selecting the best of the corresponding schedules assures a convergence towards optimality [1]. When 'enough' schedules are obtained, the best one is selected as the schedule for the day. Since a schedule can be generated in a fraction of a second, hundreds of schedules can be obtained in just a few minutes of computer time.

The idea of a 'best' schedule implies that each schedule can be evaluated. This in turn implies that there exist scheduling goals. Unfortunately, no formal enumeration of goals existed at the missile range. But, when current scheduling practices were investigated, certain criteria were found which influenced the choice of jobs which would be scheduled and those which would be deferred. These are summarized and then given precise definitions as in a, b and c, below:

a. Absolute priority (a number between one and four assigned to the job by a controlling authority to represent the relative importance of the job);

b. The number of days waited (which in effect increases the priority since the more days a job has been delayed, the more important it is to have it run); and

Vol. 4
1976

c. Project readiness (the probability—updated periodically—that the job will run if scheduled. This information can be obtained from historical data on the performance of the contractor.)

We were able to combine these factors into a single expression from questions which drew on the experience of the current scheduling staff. First of all, the priority was of particular importance. Furthermore, the importance was not linear with the priority value. Instead a quadratic loss seemed to approximate the experience of the staff reasonably well. But, the past performance of different contractors was also a factor; this was measured by the readiness value. A multiplicative factor of the quadratic value of the priority with the readiness seemed to be satisfactory. Finally, if a job could not be scheduled, all agreed that the job value should be upgraded the following day. The following job value function evolved:

$$\text{Job value} = [(\text{priority})^2 \times (\text{readiness})] + (\text{days waited}).$$

Needless to say, this function was determined somewhat arbitrarily and, as with most functions which attempt to measure subjectivity, only approximates the actual goals of the range. However, to date, a better function has not been found, although different forms have been and will continue to be examined.

With this definition of job value, the schedule payoff can be defined as the sum of the job values of those jobs that are scheduled. Thus, the quality of a given schedule is somewhat dependent on the number of jobs that are scheduled to be run in an 8 hr work day.

One limitation of Monte Carlo methods was discovered and eliminated early. The random generation of schedules can be quite inefficient since many of the random sequences are often far from optimal. Therefore, much computation effort is spent in generating sequences of mediocre value. However, neighborhood search techniques, as discussed by Petersen [9] and by Reiter and Sherman [12], tend to increase the efficiency of Monte Carlo procedures by exhaustively searching all sequences in a 'neighborhood' of a given solution.

Green and Randolph [7] describe a method of constructing a neighborhood of a random permutation for missile range scheduling. After the schedule for a given permutation is determined and evaluated, the first term of the permutation is put at the end and each remaining term is shifted-up one position. With this new permutation a new schedule is constructed and evaluated. This shifting is continued through the complete permutation and the best of these schedule payoffs is used as the local maximum.

As can be seen, with N-jobs this neighborhood structure results in a local maximum that is a random variable incorporating information from, essentially, N-schedules. By taking advantage of the structure of the problem, it is possible to generate this one random local maximum in approximately two-thirds of the time it would take to generate N individual random schedules. Furthermore, as reported by Green and Randolph [7], the average of the local maxima is signi-

ficantly greater, while the standard deviation shrinks to a fraction of the values of individual schedules. Thus, the experimental evidence indicates that this neighborhood structure will reduce the total time required to find an optimum or near optimum schedule.

It should be noted that both Peterson [9] and Reiter and Sherman [12] suggest a variety of more complex neighborhood structures. Almost all of these were examined and discarded. In general, the computational time needed to find a local maximum soared but improvement in the mean and variance of the local maxima was negligible. This disappointing experience with the more complicated neighborhood structures was possibly due to the nature of the missile range scheduling problem.

Peterson [9] and many other authors recommend a variety of dispatch rules. We tried scheduling in the order suggested by priority values, by length of the job, by the number of pieces equipment used, etc. All these methods were tried both deterministically and probablistically. Our experiments all pointed to poorer results with dispatch rules than without. Apparently dispatch rules tend to prevent certain types of permutations from being generated and, thus, restrict the domain.

Vol. 4
1976

Also, Reiter and Sherman [12] suggest that if a good permutation is observed, future permutations should be chosen in the same general vicinity as the good one. Experiments with this rule were also discouraging so the idea was abandoned.

When 'enough' random local maxima are obtained, the random schedule generation can be terminated and that schedule corresponding to the best of the observed scheduled payoffs can be used for scheduling the day's mission requests. The problem, of course, is to determine how much is 'enough'; in other words, when to stop the random generation of schedules.

STOPPING RULES

Let X_1, X_2,. . . denote the random variable of the payoffs associated with generating successive local maxima by a Monte Carlo sampling process. For the present, assume that each schedule payoff is an integer and that the objective of the scheduling problem is to find a schedule for which the payoff is maximized. Furthermore, without loss of generality, assume that all payoffs are positive and bounded above by the known integer l. In missile range scheduling, this upper bound is obtained by assuming that all jobs requested for a day are scheduled. Also, let y_n denote the maximum of the observations $k_1,...,k_n$, obtained from generating the first n local maxima; that is, $y_n = \max(k_1,...,k_n)$, where $k_1,...,k_n$ are the sample observations of the random variable $X_1,...,X_n$. This assumes that sampling will be with recall.

Since the payoffs are positive integers, the probability function for each local

maximum is the multinomial characterized by $P(X=k)=p(k),k=1,...,l$. When the values of $p(k)$ are known, then, as Chow and Robbins [4] have shown, the optimal stopping rule is obtained by calculating the expected increase in gross payoff associated with generating another local maximum,

$$T(y_n) = \sum_{k=y_n}^{l} (k - y_n)\, p(k),$$

and comparing this with the relative cost, c, of generating a single local maximum on the computer. That is, if $T(y_n) > c$, continue for another observation; otherwise stop. This procedure is called a 'myopic' because at each stage the experimenter does not consider the possible influence of observations beyond the succeeding one when making his decision. For sampling with a known distribution the myopic procedure is optimal.

Unfortunately, for scheduling problems the values of $P(X=k)=p(k),k=1,...,l$ are rarely available. However, even though these probabilities are not known, it is possible to obtain estimates of them through a Bayesian analysis. These Bayesian estimates of $p(k)$ are substituted into the above myopic stopping rule function to obtain a Bayesian stopping rule for multinomial observations. As shown in [10] this is given as

$$T_B(y_n) = (m/(m+n)) \sum_{k=y_n}^{l} (k - y_n)\, p_o(k),$$

where the $p_0(k)$ are initial prior estimates of the probabilities and $m>0$ is a confidence coefficient. The initial probabilities, $p_0(k)$, reflect the experimenter's beliefs, prior to the start of the computer run, that a local maximum of value k will be observed. The confidence coefficient, m, is roughly analogous to the sample size necessary to obtain the quality of estimates as given by the prior probabilities. If m is close to zero, this means that the experimenter has little confidence in his prior probabilities and will look to the observations for additional information on the characteristics of the probabilities. As m becomes larger, the experimenter has more and more confidence in his prior probabilities.

In missile range scheduling the mission payoffs will not be integers. Furthermore, the number of different possible values of the payoffs is large and also the values of these payoffs are unknown. Thus, for missile range problems the schedule payoffs can be considered essentially as continuous random variables and the integer stopping rule above must be modified.

We experimented with many possible prior distributions. Initially, we were attracted to the uniform distribution since it seemed to reflect our general ignorance of the prior probabilities of the process. Indeed, since each day presents a new set of jobs, we thought it likely that there would be no carry-over from one day to the next.

However, when we used the scheduling algorithm to generate several thousand local maxima for each of several different problems, we, like Heller [8], found that the distribution of local maxima tends to be very close to normal. Since each schedule payoff is formed from the sum of a random number of job payoffs, we speculate that some kind of central limit theorem is at work here. Based on these experiments we decided that a normal prior distribution was more appropriate than any other distribution. Furthermore, after a little experience, personnel at the missile range have developed a skill at estimating the mean and standard deviation of this normal prior distribution from the number of jobs available, their types, the amount of resources available and so forth. However, we keep a running mean and variance in our printout for use in training.

In the stopping rule it is assumed that the expected improvement in schedule payoffs, $T(y_n)$, could be compared with the cost of running a computer. The cost of running a computer can be estimated fairly accurately in terms of dollars and cents, although there are some uncertainties, especially for a government installation. On the other hand, the value of the schedule payoffs, $T(y_n)$, is very difficult to translate to units of dollars.

The major factor in the expression for schedule payoff is the quadratic value of the priority. But what is the meaning of that? We spent considerable time discussing this concept with people involved with scheduling and with people involved with the conduct of tests on the range. Near the top of the list of factors which could be assessed in terms of cost were, of course, the amount and types of equipment required and the number of personnel involved. But these had to be balanced against such intangibles as coordination of far-flung units, working schedules, prestige, etc. Nor could we fail to take into account the fact that at a certain point the computer might be tied-up interminably with little possibility of improving the best schedule available so far. Eventually a number was produced which has since proved satisfactory, a sort of quasi-dollar, quasi-utility measure which we represent by the letter c. The actual numerical value of c is not important for this paper.

SCHEDULE CONSTRUCTION

The computer program which was developed to construct the schedules was large and involved. Initially, we partitioned the program into three parts:

1. *Data file construction.* All information concerning a job is listed in booklets. Copies of these booklets are distributed to many offices around the missile range. Since one complete set of booklets requires five file cabinets, literally dozens of file cabinets are used merely for storage of this information. Furthermore, any changes in job requirements initiates extensive effort for modifications of these files. Thus, the construction of data files on magnetic tape, solely to

store job requirements, is desirable, especially since the many computer terminals installed throughout the missile range assure easy access to this information.

2. *The scheduling algorithm.* The process of building a conflict free schedule based on a random permutation was not a simple matter. The core of the program was a conflict-table. We divided the work day into 96 5-min periods, assuming that a 5-min period formed a sufficiently fine time-grid for scheduling purposes. We took advantage of the bit-manipulation characteristics of the UNIVAC 1108, so this table required but little computer storage, even though up to three hundred different resources had to be considered.

The program schedules the jobs (according to the permutation) by checking the conflict table for each required resource starting at the beginning of the work day. If a time–space of sufficient size is found in the conflict-table in which no resource for the particular job under test has been scheduled, the job is scheduled for that time–space by eliminating the time slot needed by each resource from further consideration. If no such space exists, the computer program records the job as unscheduled. By adding the job-payoffs for the scheduled jobs, the schedule-payoff is obtained.

Also, by dividing the set of jobs into a few (three or four) subsets, where any job in a given subset has a resource in common with at least one other job in the subset, but none in common with any job outside the subset, considerable checking can be eliminated, and the time required for the generation of schedules for local maxima can be decreased. That is, when the first job of a permutation is moved to the end, only those jobs in its subset need to be rescheduled.

3. *The output.* After a final schedule has been determined, it is necessary to transmit the relevant information to the contractors, machine operators, dispatchers, the range safety office, etc. This requires several different kinds of output. For example, a radar operator needs a list of the jobs which will use his radar and their times. The contractor needs only the information relevant to his job, namely the T-time and the time each resource is to be ready. The dispatcher needs, essentially, a complete Gantt chart.

Our initial strategy was to concentrate our efforts on part 2, the scheduling algorithm. We reasoned that a feasible scheduling algorithm was needed or there would be no point in working on the output format. Furthermore, even though the data files would be useful for the missile range without computerized scheduling, it became evident that only a working scheduling algorithm could provide sufficient justification to management for the large outlay of time and expense necessary to construct such data files.

It turns out that this partial effort was ill advised and almost resulted in a severe set-back to the project. Since the data files were not ready, we used fictitious data for the scheduling program. After several months we were ready to demonstrate our program. But our output—perfectly clear to us—was not clear to the officials who viewed the demonstration. They did not recognize our jobs (since they and the data were fictitious) nor did they understand the

form of the output. Many explanations and several consultations were required (which occupied jealously alloted time) to resolve these difficulties. After some arbitration it was decided that we should engage in a data-gathering project whose purpose was to define types of output and determine suitable formats which would correspond to various kinds of requirements. This, perforce, included the charts and graphs beloved of management. After we had gathered sufficient knowledge to lay-out a variety of provisional output formats, we tested our assumptions on a few selected individuals. We received a number of suggestions which we incorporated into our output programs, introduced some real data obtained by hand from the file cabinets, and prepared another presentation. The subsequent reaction was far more favorable.

At present, our scheduling routine is in good shape and the output format seems to be satisfactory. We are using the available data files plus some hand collection of data to provide schedules for 1 or 2 days a week. There are several advantages to this lethargic conversion process. For example, it introduces personnel to the routine gradually and does not eliminate people overnight. It provides a test and break-in period for the routine, as well as for the operators, in an actual production situation. This, in turn, provides feedback to the scheduling program for inclusion or correction of procedures that were overlooked. The tedious job of building the data files is on-going and may take another year or so to complete, so that implementation is expected to take a fairly long time.

Vol. 4
1976

EXAMPLE OF A SCHEDULE

Each day's schedule request uses a code for identification called the 'Job OD', a five-digit number with up to three letters. Using these codes, the permanent data files are searched for the route sheets corresponding to the requested jobs.

The data from the route sheet, transferred into high speed core storage, are used by the scheduling algorithm to generate a series of random local maxima. An example of the form used in listing the local maxima is shown in Table 2. Note that $T(y_n)$, the expected improvement, is arbitrarily forced to remain at $1+c$ for the first six trials, where c is relative cost value. This accumulates sufficient data so that the statistical parameters can be estimated. Also note that the local maximum numbered 11 was retained although the expected improvement did not drop below the threshold value of 0.00636 until sample number 12.

The jobs in the final schedule are indexed according to the time at which they are to be run. Table 3 illustrates the indexed schedule summary resulting from schedule 11 of Table 2.

Table 1 illustrates the scheduling of job 81321E according to schedule 11 of Table 2. The words 'prime' and 'nonprime' could be replaced by equivalent terms 'necessary' and 'desirable'. The list shows the time window for each

TABLE 2. CALCULATIONS FOR THE STOPPING RULE

Largest possible expected schedule payoff is 203·450.
The confidence coefficient is: 1·3333.
The value of the relative cost is: 0·00636.

N	Schedule payoff	Largest payoff so far	Expected improvement	Mean	Standard deviation
1	62·409970	62·409970	1·006360	62·410000	0·000000
2	71·159970	71·159970	1·006350	66·785000	4·375000
3	57·219970	71·159970	1·006360	63·596600	5·752508
4	72·369970	72·369970	1·006350	65·789909	6·265034
5	70·889960	72·369970	1·006360	66·809990	5·963398
6	73·339960	73·339960	1·006360	67·898300	5·963007
7	73·419960	73·339960	0·103791	68·544290	5·742938
8	62·219970	73·339960	0·072758	67·753690	5·764833
9	62·249960	73·339960	0·051852	67·142190	5·703721
10	64·999960	73·339960	0·037717	66·927990	5·449056
11	81·169960	81·169960	0·006803	68·222700	6·614840
12	69·609970	81·169960	0·005073	68·338300	6·334823

TABLE 3. LIST OF SCHEDULED MISSIONS

Index	SRVC	OD Num	Short title	*T*-Time	Radar	Telem	Optics	Elec-S
AA	A	04904C	Grande Prix	715	Yes	No	No	Yes
AB	F	81324G	Torino	720	Yes	No	Yes	No
AC	N	70721	El Camino	730	Yes	Yes	Yes	Yes
AD	R	70724	Roadmaster	730	Yes	Yes	Yes	Yes
AE	A	40722A	Travelall	730	Yes	Yes	Yes	No
AF	F	81310A	New Yorker	800	Yes	Yes	Yes	No
AG	N	45708B	Hupmobile	805	Yes	No	Yes	No
AH	R	81324H	Rambler	825	Yes	Yes	Yes	No
AJ	A	81310B	Tempest	825	No	Yes	No	No
AK	F	81321E	Bonneville	910	Yes	Yes	Yes	Yes
AL	N	81317V	Le Mans	930	Yes	No	Yes	No
BA	R	81319A	Firebird	950	Yes	No	Yes	No
BB	A	80204	Catalina	1000	Yes	Yes	Yes	No
BC	F	80206	Riviera	1000	Yes	Yes	Yes	Yes
BD	N	81317W	Ambassador	1020	Yes	No	Yes	No
BE	R	29088	Monte Carlo	1100	Yes	Yes	Yes	Yes
BF	A	32004C	Imperial	1100	Yes	Yes	Yes	No
BG	F	45709C	Malibu	1130	Yes	No	Yes	Yes
BH	N	80202A	Le Sabre	1140	Yes	No	Yes	No
BJ	R	80202B	Electra	1200	Yes	Yes	No	No
BK	A	80722G	Monsa	1200	Yes	No	Yes	Yes
BL	F	40722F	Cutlass	1215	Yes	No	Yes	No
CA	N	03904C	Pinto	1300	Yes	No	Yes	No
CB	R	45708C	Pierce Arrow	1300	Yes	Yes	Yes	Yes
CC	A	40722D	Valient	1300	Yes	No	Yes	No
CD	F	45708A	Marquis	1330	Yes	Yes	Yes	Yes
CE	N	70720	Maverick	1330	Yes	Yes	Yes	No
CF	R	80722A	Skylark	1405	No	Yes	No	Yes
CG	A	80722F	Commando	1410	Yes	No	Yes	Yes
CH	F	80722F	President	1415	Yes	Yes	Yes	No
CJ	N	00608T	Javeline	1450	Yes	Yes	Yes	Yes
CK	R	40722B	Impala	1500	No	Yes	Yes	No

resource and the required quantity of that type of resource needed. The '*T*-TIME' is 'all-systems-go' time. Certain resources may require a warm-up time before '*T*-TIME' and a 'turn-around' or 'tear-down' time after completion. This information is indicated by the 'Beginning Time' and the 'End Time' in Table 1. Finally, Table 4 illustrates a truncated listing of unscheduled jobs and the identification of the resource conflicts with the scheduled jobs.

TABLE 4. RESOURCE CONFLICT IDENTIFICATION FOR UNSCHEDULED MISSIONS

Unscheduled mission		Resource conflicts Among scheduled missions			
O D Number	Short title	Index	O D Number	Short title	Resource
81322H	Galaxy	AA	04904C	Grand Prix	R114
		AB	81324G	Torino	R125
					R123
					R122
		AC	70721	El Camino	R125
					R122
		AD	70724	Roadmaster	F109
					R113
					R114
					R125
					J56
					D89
		AE	40722A	Travelall	R114
					R125
					R112
					R123
					R122
		AF	81310A	New Yorker	R125
					R123
		AG	45708B	Hupmobile	R112
		AH	81324H	Rambler	R114
					R125
					T866
					R122
		AK	81321E	Bonneville	J10
					R125
					T866
					R112
					R123
					R122
		AL	81317V	Le Mans	R114
		BA	81319A	Firebird	R123
		BB	80204	Catalina	R125
		BC	80206	Riviera	R113
					R114
					R125
					R112
		BD	81317A	Ambassador	R114
					R115

Vol. 4
1976

The computer time for this example happened to be under 1 min on the central processor unit. For production runs this will tend to be higher because the threshold value is lower. We chose 0.00636 for the value of c to avoid a lengthy computer printout for this paper. A typical time is 5 to 10 min, usually run during low priority hours such as the middle of the night or on weekends. On the other hand, if, as sometimes happens, all jobs requested can be fitted into the work day, an expected improvement value of zero is soon obtained. The amount of computer time then becomes very small, usually only a few seconds.

CONCLUSIONS

Although the automated computerized Monte Carlo scheduling algorithm may not exhibit the elegance of some other approaches, it nevertheless produces schedules for real problems, it exhibits a convergence to optimality, it operates in a predictable fashion within a probabilistic and statistical framework, and it is sufficiently flexible to handle the variety of scheduling problems that occur at a missile range. Because of its generality, the algorithm could probably be adapted to almost any scheduling problem.

Already, although limited due to incomplete data files, the automated Monte Carlo scheduling algorithm has proven to be of considerable value to the missile range. It has forced the construction of a long overdue computer file on the jobs and their requirements. More important, it is producing schedules that are at least as good as, and usually better than, those generated manually. Furthermore, the computer-generated schedules are produced in minutes at a fraction of the cost, rather than the days needed by a couple of dozen people. In addition, the computer-generated schedules are conflict free, eliminating the confusion and harried rescheduling efforts that attendant conflicts create.

There are, perforce, several problems yet to overcome. Careful and gradual implementation of the algorithm enables us to solve many of the internal programming problems as they arise. Also, we are still examining the payoff function. In its present form, as mentioned above, it is reasonably successful but does not entirely reflect the expectations of the missile range in that the algorithm doesn't consider overtime. But authorized overtime represents a kind of negative payoff, which is, subjectively, justifiable at times for some jobs but not at other times or for other jobs. A variety of imponderables such as future workload (i.e. could a job be scheduled in a day or two because of relatively light requirements?), season of the year (vacation, for example), weather, both past and anticipated future, etc. affect scheduling practices. These problems will require considerable further thought and judgement.

REFERENCES

1. BEENHAKKER HL (1963) Mathematical analysis of facility—commodity scheduling problems. *Int. J. Prod. Res.* **2,** 313–321.
2. BROOKS G and WHITE CR (1965) An algorithm for finding optimal or near optimal solutions to the production scheduling problem. *J. Ind. Engng* **16,** 34–40.
3. BROWN JR (1970) Subductive programming and chromatic scheduling. PhD Thesis. Massachusetts Institute of Technology.
4. CHOW YS and ROBBINS H (1963) On optimal stopping rules. *Z. Warscheinlichkeitstheorie* **3,** 33–49.
5. FENEMA FF (1965) Scheduling against a probabilistic demand in a heuristic environment. PhD Thesis. Oklahoma State University.
6. GRIFFLER B and THOMPSON GL (1960) Algorithms for solving production scheduling problems. *Ops Res.* **8,** 487–503.
7. GREEN R and RANDOLPH PH (1973) How to beat Monte Carlo: an improved scheduling technique. *Trans. Eighteenth Conf. Army Mathematicians.* ARO-D Report 73-1, Office of the Chief of Research and Developments, Army Research Office, Durham.
8. HELLER Jack (1960) Some numerical experiments for an $M \times J$ flow shop and its decision theoretical aspects. *Ops Res.* **8,** 178–184.
9. PETERSON CC (1973) Solving sequencing problems through reordering operations. *AIIE Trans.* **5,** 68–73.
10. RANDOLPH PH, SWINSON GW and ELLINGSEN C (1973) Stopping rules for sequencing problems. *Ops. Res.* **21,** 1309–1315.
11. RANDOLPH PH (1973) Stopping rules for sequencing problems with particular reference to missile range scheduling. *Proceedings of the Eighteenth Conference on Design of Experiments in Army Research, Development, and Testing.* Report Number 73–2, US Army Research Office, 233–8.
12. REITER S and SHERMAN G (1965) Discrete optimizing *J. Soc. Ind. Appl. Math.* **13,** 864–889.
13. WAGNER HM (1959) An integer programming model for machine scheduling. *Nav. Res. Logist. Q.* **6,** 131–140.

ADDRESS FOR CORRESPONDENCE: *P Randolph, Federal Energy Administration, Room 4519 New Post Office Building, 12th and Pennsylvania, Washington DC 20461, USA.*

Vol. 4
1976

BLANK
PAGE

Vol
19

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

An Approach to Stochastic Programming for Medium Term Planning[1]

ROBERT G DYSON

GILLIAN SWAITHES

Received July 1975; in revised form March]1976)

A number of methods of obtaining the distribution of the optimum of the 'wait and see' stochastic programming model have been proposed, but computational experience for these is currently limited to the solution of small problems. The purpose of this paper is to discuss the role of the 'wait and see' model in planning, and to propose a method of analysis based on the minimax and maximax decision criteria. The approach requires the solution of a special class of non-linear programming problems. Computational results to date suggest that it will be possible to analyse practically sized problems in this way.

Vol. 4
1976

INTRODUCTION

MATHEMATICAL programming models can be classified by their usage into 'here and now' models and 'wait and see'[2] models. In 'here and now' models, the set of values of the decision variables is a proposal for action evaluated by the objective function, and in a stochastic model some of the values must be determined before the outcomes of the random events are known. Examples of situations where 'here and now' models have been used include blending problems and air crew scheduling.

The purpose of the 'wait and see' model is to compare the effectiveness of alternative proposed systems. For example, we may have a model of an existing production or distribution system and wish to evaluate proposed changes to the system. The changes might include the purchase of a new piece of equipment, a

[1] This research was supported by SRC.

[2] The expression 'wait and see' was first used by Madansky [15] and the model was first studied by Tintner [17].

change in marketing policy leading to different demand levels, or alternative supply contracts for raw materials. Such changes can be evaluated by modifying the appropriate parameters of the model and re-running. The change in the value of the objective function (which may be contribution to profits and overheads) is then a measure of the effectiveness of a proposed change; Gupta and Rosenhead [11] use a transportation model in this way to analyse a distribution situation.

A consequence of using mathematical programming models in this manner is that the evaluation must necessarily be carried out some time in advance of implementation (say up to 2 years). The parameters of the model are therefore estimated in advance, and may exhibit a high degree of uncertainty. In the stochastic case it is assumed that the levels of the decision variables need be set only after the outcomes of the random parameters are known. The decision variables and hence the value of the objective function will thus be random variables. It might therefore be appropriate to seek the distribution of the optimum values so that different systems may be compared via the expected value and variance of the objective function.

One method of obtaining the distribution of the objective function of the 'wait and see' model would be to use simulation. This approach is, however, very time-consuming, and consequently researchers have sought analytical methods. Bereanu and Peeters [4] have given an analytical solution to the problem where the uncertainty is confined to the RHS coefficients of a linear program and each stochastic parameter can be described by a linear function of a single random variable by the use of parametric programming. For the extension to the case of multiple random variables, Bereanu [3] has developed approximate methods based on cartesian integration for solving the resultant multiple integrals. At present it is possible to solve problems containing a few random variables in this way. Ewbank, Foote and Kumin [8] also adopt a multiple integration approach, and have proposed a transformation which simplifies the regions of integration and often simplifies the integration itself. Dempster [5] proposes a method of approximating the distribution of the optimum using quantile arithmetic. There is little computational experience of these methods to date.

To summarize, we are considering the use of 'wait and see' problems in planning where a deterministic model is an inadequate representation. A simulation approach may require a prohibitive amount of computation, and analytical methods of determining the distribution of the objective function are currently available only for small problems. In the following sections we propose a method of analysis based on the minimax and maximax decision criteria.[3] The method aims to handle the uncertainty better than is possible

[3] The methods of solution of the 'wait and see' problem described here, would provide minimax solutions to 'here and now' problems if required.

using a deterministic model, whilst avoiding the severe computational problems that arise in seeking the distribution of the optimum.

THE MINIMAX APPROACH TO THE 'WAIT AND SEE' PROBLEM

We are concerned here with using mathematical programming models to evaluate proposed changes to an object system, and are making the assumption, characteristic of 'wait and see' models, that the levels of the decision variables need be set only after the outcomes of the random parameters are known.

If we use a deterministic model of the situation under study, obtained by replacing the random parameters by their most likely values, the solution determines the most likely value of the objective function. Our proposal is that minimax solution (occurring when the random parameters take on their pessimistic values) and the maximax solution (optimistic parameter values) are also obtained, thus providing bounds on the value of the objective function for use as additional measures for evaluation. In particular, the minimax solution gives a good indication of the robustness of a proposed system in the face of uncertainty. The different measures may of course give conflicting advice about alternative proposals and the eventual choice of a proposal might well depend on the decision maker's attitude to risk.

Vol. 4
1976

The minimax and maximax solutions may also assist in formulating proposed changes to a system. The shadow prices of the solution to the deterministic model indicate the marginal values of increasing the various resources, and thus can suggest beneficial changes to the system such as the alleviation of production bottlenecks. In stochastic situations a proposal formulated as a result of a high shadow price in the deterministic solution may turn out to be uneconomic. If the minimax and maximax solutions are also obtained, then a proposal to increase a resource having a high marginal value in each case should prove attractive in a wider variety of emerging situations.

In simple examples, an indication of the robustness of a proposal to the uncertainty could be obtained using sensitivity and post-optimal analysis but these are 'one parameter at a time' techniques and are increasingly ineffective as the magnitude and number of the uncertain parameters increases. It is in these more uncertain situations that the minimax and maximax solutions should prove most valuable.

Plausible regions

In the previous section we assumed knowledge of the minimax and maximax values of the uncertain parameters. For a single stochastic parameter the minimax and maximax values would be our pessimistic and optimistic estimates of the parameter value. These estimates bound a region which we will call the

'plausible region'. Our assumption here is that this region will include, with probability $(1 - \alpha)$, the random variable describing the stochastic parameter. We are thus suggesting that there will be small chances of the actual parameter value being greater than the maximax value or less than the minimax value, but we are prepared to discount these events as being implausible. This approach is on the lines of Guilbaud [10, p. 79] who states 'Man acts as if events with very small probability ought not to happen at all'.

If we have multiple stochastic parameters, determining the minimax and maximax values of these parameters may not be straightforward. The likelihood of all the parameters taking on their pessimistic values simultaneously will be very small, unless the parameters exhibit high positive correlation, (or are generated by the joint exponential distribution!), and we may consider such an event implausible. Consequently it is not obvious in the multivariate case which set of parameter values would lead to the minimax solution. The problem can be resolved by carrying over the plausible region concept to the multivariate case. Hence we wish to define a region containing the stochastic parameters with joint probability $(1 - \alpha)$. In general there will be an infinity of such regions, but we propose that a region bounded by a contour of constant likelihood is appropriate. Its attraction is that all points within the region will have a higher likelihood or frequency than those excluded, and it will consequently be the smallest $1 - \alpha$ region. For the multivariate normal distribution, for example, the plausible region will be defined by the hyper-ellipsoid

$$(\mathbf{Z} - \boldsymbol{\mu})\, \mathbf{V}^{-1}\, (\mathbf{Z} - \boldsymbol{\mu}) \leqslant K$$

where $\mathbf{Z}$ is a vector of stochastic parameters, $\boldsymbol{\mu}$ is a vector of means, $\mathbf{V}$ is the covariance matrix and K is a constant term defining the size of the region which is related to α. In general the plausible region, R, will be represented by a non-linear constraint of the form:

$$R\colon G(\mathbf{Z}) \leqslant K,$$

and minimax and maximax solutions can be obtained by incorporating the constraint into the mathematical program. The resultant minimax and maximax problems are discussed in the following section.

To complete the specification of R it is necessary to determine K. In some specific instances it may be possible to determine K analytically (Johnson and Kotz [12], p. 87), but in general it will be necessary to resort to numerical methods (e.g. Monte Carlo methods, Stroud [16]). Alternatively it may be possible to obtain directly sets of parameter values which are just plausible, and construct R on these points.

THE MINIMAX AND MAXIMAX PROBLEMS

In this section we shall confine our attention to situations where only RHS coefficients in less than or equal to inequalities are stochastic, i.e.

$$
\begin{array}{lllll}
 & x_0 & = \mathbf{c}'\mathbf{x} & & \\
\text{subject to} & \mathbf{A}\mathbf{x} & \leqslant \mathbf{b} & & \text{the deterministic constraints,} \\
 & \mathbf{A}_1\mathbf{x} & \leqslant \mathbf{Z} & & \text{the stochastic right hand sides,} \\
 & G(\mathbf{Z}) & \leqslant K & & \text{the plausible region.} \\
 & \mathbf{x},\mathbf{Z} & \geqslant \mathbf{0}. & &
\end{array}
$$

The maximax problem requires us to maximize x_0 with respect to $\mathbf{x}$ and $\mathbf{Z}$ and is a convex non-linear programming problem which is linear other than for a single non-linear constraint (provided that the plausible region is convex). There are a variety of approaches to solving non-linear programs, and we have been using the method of supporting hyper-planes due to Kelley [13].

For the minimax solution we must maximize x_0 with respect to $\mathbf{x}$ for given $\mathbf{Z}$, and then minimize with respect to $\mathbf{Z}$. The nature of the objective function causes the problem to be non-convex. (This is also the case for the linear minimax problem, see Falk [9]). For different sets of values of $\mathbf{Z}$, different bases $\mathbf{x}$ will be optimal, and in fact the plausible region will be spanned by a number of non overlapping convex regions, each region corresponding to a different optimal basis $\mathbf{x}$. Such a convex region of parameter values has been called a decision region by Bereanu [2]. Dyson and Swaithes [6] show that a global minimax solution occurs on a subset of the plausible region, and the values of $\mathbf{Z}$ corresponding to a global minimax solution can be sought by minimizing a linear objective function over each decision region included in this subset.

An algorithm for the maximax problem and a local algorithm for the minimax problem, where the joint distribution of stochastic RHS parameters is multivariate normal, have been implemented on an ICL 1906 A using the Land and Powell suite of Fortran subroutines [14]. Experiments are currently being carried out with a model of a cannery factory. The model has 98 constraints, 190 variables and up to 20 stochastic parameters. Run times are from 2–3 min.

Details of these algorithms appear in Dyson and Swaithes [7]. A global minimax algorithm for stochastic RHS has been specified by Dyson and Swaithes [6] and extensions to include stochastic objective-function coefficients are also outlined. There should be no difficulty modifying the algorithms for different distributions, provided that the resultant plausible regions are convex. This would not be the case, for example for a bi-modal distribution as the plausible region will be disjoint.

Vol. 4
1976

CONCLUSIONS

The main purpose of this paper was to propose a method of analysing 'wait and see' stochastic programming problems based on the minimax and maximax decision criteria. We have discussed the purpose of 'wait and see' models and shown how the minimax approach can aid decision making in certain situations. We have introduced the concept of a plausible region of parameters and argued that such a region should be bounded by a contour of constant likelihood. The maximax problem is shown to be a non-linear convex programming problem and the minimax problem to be non-convex.

In the minimax approach, the experimental algorithms for maximax and local minimax solutions are capable of solving medium sized problems as indicated previously. This has been achieved without investigating alternative non-linear programming algorithms to find the one most suited to this class of non-linear problem. In particular, van de Panne [18] has suggested converting a program with a quadratic constraint into a parametric quadratic program, and our investigations suggest that this conversion could be applied to Beale's algorithm [1] rather than to the quadratic programming method based on the Kuhn–Tucker conditions. For problems with a quadratic plausible region, therefore, the prospects of finding global minimax solutions for programs somewhat larger than those currently under development seem good. More generally, the approach seems capable of analysing problems measurably greater than the distribution approaches of Bereanu, Ewbank, Foote and Kumin, and Dempster.

The minimax approach is a relatively simple method of analysing stochastic 'wait and see' problems. Computational experience to date suggests that it will be possible to develop the approach to solve a variety of practical planning problems, and thus may have wider application than distribution approaches.

ACKNOWLEDGEMENTS

The authors wish to thank EML Beale, H Boothroyd and a referee for their helpful comments at various stages in the preparation of this paper.

REFERENCES

1. Beale EML (1955) On minimizing a convex function subject to linear inequalities. *Jl R. Statist. Soc.* **A17**, 173–184.
2. Bereanu B (1964) Regions de decision et repartition de l'optimum dans la programmation lineaire. *C.R. Acad. Sci., Paris* **259**, 1383–1386.
3. Bereanu B (1971) The distribution problem in stochastic linear programming. The cartesian integration method. Report No. 7103, Center of Mathematical Statistics, Bucharest.

4. BEREANU B and PEETERS G (1968) A 'wait and see' problem in stochastic linear programming. An experimental computer code. Core Discussion Paper No. 6815.
5. DEMPSTER M (1974) An application of quantile arithmetic to the distribution problem of stochastic linear programming. *Bull. I.M.A.* **10**(5/6), 184–194.
6. DYSON RG and SWAITHES G (1975) A global algorithm for minimax solutions to a stochastic programming problem. C.I.E.B.R. Research Paper No. 61. University of Warwick, Coventry.
7. DYSON RG and SWAITHES G (1975) Fortran codes for a global maximax and a local minimax solution to a stochastic programming problem. C.I.E.B.R. Research Paper No. 62. University of Warwick, Coventry.
8. EWBANK J, FOOTE B and KUMIN H (1974) A method for the solution of the distribution problem of stochastic linear programming. *SIAM Jl Appl. Math.* **26**(2), 225–238.
9. FALK JE (1973) A linear max-min problem. *Mathl. Prog.* **5**(2), 169–188.
10. GUILBAUD GT (1959) *What is Cybernetics*, p. 79. Heinemann, London.
11. GUPTA SK and ROSENHEAD JV (1968) Robustness in sequential investment decisions. *Mgmt Sci.* **15**(2), 18–29.
12. JOHNSON NL and KOTZ S (1972) *Distributions in Statistics: Continuous Multivariate Distributions*, p. 87. Wiley, New York.
13. KELLEY JE (1960) The cutting plane method for solving convex programs. *Jl SIAM* **8**(4), 703–712.
14. LAND AH and POWELL S (1973) *Fortran Codes for Mathematical Programming*. Wiley, New York.
15. MADANSKY A (1960) Inequalities for stochastic linear programming problems. *Mgmt Sci.* **6**, 197–204.
16. STROUD AH (1971) *Approximate Calculations of Multiple Integrals*. Prentice-Hall, New Jersey.
17. TINTNER G (1955) Stochastic linear programming with applications to agricultural economics, In *Proceedings of Second Symposium in Linear Programming*, Vol. 1, pp. 197–228. Washington D.C. National Bureau of Standards.
18. VAN DE PANNE C (1966) Programming with a quadratic constraint. *Mgmt Sci.* **12**(11), 798–815.

ADDRESS FOR CORRESPONDENCE: *Dr RG Dyson, School of Industrial and Business Studies, University of Warwick, Coventry, Warwickshire CV4 7AL, UK.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Memoranda

Contractual Sharing Rules

THIS NOTE will build on the contractual incentives model [1, 2], by considering a special class of incentive functions that we denote as contractual sharing rules. An incentive contract is characterized by a variable incentive function that consists of a variable dollar amount depending on the criteria for evaluation. In addition to the variable incentive there is a fixed payment. One person, the principal, specifies an incentive function, so as to motivate another person, the agent, to accept a contract and subsequently make appropriate decisions.

In the contractual incentives model the evaluation criteria are related to the agent's decision. However, in this paper, we shall restrict ourselves to situations where the agent does not make decisions subsequent to the contractual agreement. We shall use the term *sharing rule* to denote this restricted use of the incentive function.

In the sharing rule model the principal owns a lottery and, for one reason or another, he wants the agent to participate (possible sharing losses as well as gains and thus reducing risk). A number of environments are compatible with such a scenario. For example, the principal may, in the given lottery (or investment) need managerial assistance, additional capital, a diversification of risk, etc. So we are assuming that, from the principal's viewpoint, the agent's participation is preferable to the situation that would exist without the agent's participation.

In specifying a contract to offer the agent, the principal considers possible sharing rules characterized by two dollar amounts. One is a variable amount $s(c)$, depending on the evaluation criterion c. The other is a fixed dollar amount H. Thus the agent would be paid $s(c) + H$. This form is not restrictive since H can always be set equal to zero.

After being offered the contract the agent then decides, given his other options, whether or not to accept. We shall let L represent his minimum acceptance level. Thus he will accept the contract if

$$Ev[s(c) + H] \geq L \tag{1}$$

where $v(\cdot)$ is the agent's utility function on the uncertain payment.

The principal's problem then becomes

$$\text{Max } U \equiv Eu[g - s(c) - \text{H}] \tag{2}$$
$$\text{subject to } V \equiv Ev[s(c) + H] = \text{L}$$

where $u(\cdot)$ is the principal's utility function on his net results, and g is his net return before paying the agent.

The subsequent development will no longer utilize c as a variable and as a simplifying measure we will let $S \equiv s(c)$. It will also be helpful to use z to index uncertainty, so we will let $g \equiv g(z)$ and $S \equiv S(z)$ where $z \in \{z\} \equiv Z$ and $\pi_i(z)$ is a probability distribution on the discrete set Z, [$i = a$(agent), p(principal)].

For a specific $z' \in Z$ we obtain

$$\frac{\text{d}U}{\text{d}S(z')} = -u'[G(z')]\,\pi_p(z') - \frac{\text{d}H}{\text{d}S(z')}\,\Sigma_z u'[G(z)]\,\pi_p(z) \tag{3}$$

and from the equality

$$\frac{\text{d}V}{\text{d}S(z')} = v'[R(z')]\,\pi_a(z') + \frac{\text{d}H}{\text{d}S(z')}\,\Sigma_z v'[R(z)]\,\pi_a(z) = 0 \tag{4}$$

where $\pi_p(z)$ and $\pi_a(z)$ are the respective probability distributions of the principal and agent; and

$$G(z) \equiv g(z) - S(z) - H$$

and

$$R(z) \equiv S(z) + H.$$

Vol. 4
1976

From (4) we solve for $dH/dS(z')$ and substitute into (3) to get

$$\frac{dU}{dS(z')} = D[z'][\Sigma_z u'[G(z)]\pi_p(z)]/u'[G(z')] \quad (5)$$

where

$$D(z') \equiv A(z')\,\pi_a(z') - P(z')\,\pi_p(z'), \quad (6)$$
$$P(z') \equiv u'[G(z')]/\Sigma_z[G(z)]\,\pi_p(z),$$
$$A(z') \equiv v'[R(z')]/\Sigma_z v'[R(z)]\,\pi_a(z).$$

And assuming $u(\cdot)$ is strictly increasing then the sign of $dU/dS(z')$ is solely determined by $D(z')$. Under this assumption we will now develop some results for various cases.

Under linear utility $D(z') = \pi_a(z') - \pi_p(z')$, so if $\pi_p(z) = \pi_p(z)$ for all $z \in Z$ then the principal cannot increase his utility by revising the sharing rule. By the same token if $\pi_a(z') < (>)\,\pi_p(z')$ for some $z' \in Z$ then $D(z') < (>)0$, so $S(z')$ can be decreased (increased) without limit and not reduce U.

Under strict joint risk preference (both $u(\cdot)$ and $v(\cdot)$ convex with one being strictly convex) we can find an $N(z) < \infty$ such that $A(z) > (\geq)1$ and $P(z) \leq (<)1$ for all $S(z) > N(z)$. Consequently for any $\pi_a(z') \geq \pi_p(z')$ we would have $D(z') > 0$ with the result that $S(z')$ could be increased without limit and not reduce U.

Under strict joint aversion (both $u(\cdot)$ and $v(\cdot)$ concave with one being strictly concave) then we first observe that if $\pi_a(z') = (>)0$ and $\pi_p(z') > (=)0$ then we have the unbounded condition as in the previous two cases. So for the subsequent development we will assume that $\pi_a(z) \cdot \pi_p(z) > 0$ for all $z \in Z$.

There are two additional conditions such that an unbounded solution can occur. One is $v'(\cdot)$ having a lower bound greater than zero and the other is $u'(\cdot)$ having a lower bound greater than zero. These lower bounds are necessary but not sufficient. An appropriate ratio between $\pi_a(z')$ and $\pi_p(z')$ is also required.

To have a bounded solution it is sufficient to have the following conditions:

1. strict joint risk aversion;
2. $v'(\cdot)$ and $u'(\cdot)$ with lower limits of zero.

To prove this we note that the limit of $D(z') < (>)0$ as $S(z')$ approaches ∞ $(-\infty)$. So, given the continuity, U attains a maximum value for some value of $-\infty < S(z') > \infty$.

REFERENCES

1. Berhold M (1967) An analysis of contractual incentives, PhD dissertation, University of California, Los Angeles.
   Also Working Paper No. 129, Western Management Science Institute, University of California, Los Angeles.
2. Berhold M (1971) A theory of linear profit sharing incentives. *Q. J. Econ.* LXXXV(3), 460–482.

M Berhold
(*May 1976*)

Department of Quantitative Methods
Georgia State University
University Plaza
Atlanta, Georgia 30303
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

An Approximation for the Replenishment of Perishables Problem

WHEN DEALING with the replenishment of perishable goods one is faced with the problem of determining the optimal stock level at the beginning of a time interval T so that the total cost C of leftover and runout goods over T is minimised. The demand from the stock during T is stochastic with mean $\bar{D}$ and variance σ^2. Empirical studies have shown that the normal distribution can often be assumed except for very low demand, and C is then given by [1]:

$$C = \sigma[(f + t_N F)(c_1 + c_2) - t_N]$$

with c_1 the penalty per leftover unit, c_2 the penalty per runout unit, $t_N = (Q - \bar{D})/\sigma$ where Q is the stock level at the beginning of T, f the standardized normal deviate for t_N and F the cumulative area under the normal curve up to t_N.

C is minimised when $F = 1/(1 + \epsilon)$ with $\epsilon = c_1/c_2$ (a result that also holds for all demand distributions) so that the optimal replenishment quantity $Q_{N,0}$ can be obtained as:

$$Q_{N,0} = \bar{D} + t_{N,0}\,\sigma \qquad (1)$$

where $t_{N,0}$ is the optimal value of t_N corresponding to the value $1/(1 + \epsilon)$ of F.

The normal distribution is well tabulated and the necessary calculations can be simplified further by using specially prepared Tables, such as the ones included in [2], which give $t_{N,0}$ for various values of ϵ and $C/(\sigma c_2) = (f + t_N F)(1 + \epsilon) - t_N$ for various values of ϵ and t_N. This straightforward and fairly simple analysis has, however, a disadvantage stemming from the very nature of the normal distribution in that it cannot be easily adapted to produce economies of scale. There are practical situations where this type of analysis has to be repeated several times, three typical examples being: sensitivity analyses requirements, consideration of the effects of decisions deviating from the optimal or wide range of perishable goods each with different characteristics.

The purpose of this note is to show that the triangular distribution

$$f(D) = \begin{cases} (D - \bar{D} + \sqrt{6}\,\sigma)/(6\,\sigma^2) & \text{if } \bar{D} - \sqrt{6}\,\sigma \leq D \leq \bar{D} \\ (\bar{D} + \sqrt{6}\,\sigma - D)/(6\,\sigma^2) & \text{if } \bar{D} < D \leq \bar{D} + \sqrt{6}\sigma \\ 0 & \text{otherwise} \end{cases}$$

is adaptable to produce economies of scale and yet gives good approximations for the optimal reorder quantity over a wide range of values for ϵ.

If the triangular distribution is used to provide an approximate fit to the normal demand distribution, the expressions for the optimal reorder quantity ($Q_{T,0}$) and for the total cost can be found equal to:

1. $$Q_{T,0} = \bar{D} + t_{T,0}\,\sigma = \bar{D} + (\sqrt{6} - 2\sqrt{3\epsilon/(1+\epsilon)})\,\sigma, \quad (2)$$
$$C = \sigma c_2[(1 + \epsilon)(\sqrt{6} - t_T)^3/36 + \epsilon t_T]$$
if $\epsilon \leq 1$; and

2. $$Q_{T,0} = \bar{D} + t_{T,0}\,\sigma = \bar{D} - (\sqrt{6} - 2\sqrt{3/(1+\epsilon)})\,\sigma, \quad (3)$$
$$C = \sigma c_2[(1 + \epsilon)(\sqrt{6} + t_T)^3/36 - t_T],$$
if $\epsilon > 1$, where $t_T = (Q - \bar{D})/\sigma$ and $t_{T,0} = (Q_{T,0} - \bar{D})/\sigma$.

The error incurred in using (2) or (3) instead of (1) for the determination of the optimal replenishment quantity is equal to:

$$Q_{T,0} - Q_{N,0} = (t_{T,0} - t_{N,0})\,\frac{\sigma}{\bar{D}}\,\bar{D} = \alpha\,\bar{D}.$$

Table 1 gives values of $100\,\alpha$ for $\epsilon = 0{\cdot}02 \ldots 50$ and $\sigma/\bar{D} = 0{\cdot}05 \ldots 0{\cdot}40$ in steps of $0{\cdot}05$.

It is interesting to note that an error of the same magnitude can be incurred if, in using (1), $\bar{D}$ is overestimated by 100α per cent. Judging from what one would expect as a reasonable error in estimating $\bar{D}$, the values

Vol. 4
1976

TABLE 1. VALUES OF 100α

ϵ \ $\sigma/\bar{D}$	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40
0·02	−0·5	−0·9	−1·4	−1·9	−2·4	−2·9	−3·4	−3·9
0·03	−0·2	−0·4	−0·5	−0·7	−0·9	−1·1	−1·3	−1·4
0·04	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
0·06	0·2	0·4	0·6	0·8	1·0	1·2	1·4	1·6
0·08	0·3	0·6	0·9	1·2	1·5	1·8	2·1	2·4
0·10	0·3	0·7	1·0	1·4	1·7	2·1	2·4	2·8
0·15	0·4	0·7	1·1	1·5	1·8	2·2	2·6	3·0
0·20	0·3	0·7	1·0	1·4	1·7	2·0	2·4	2·7
0·30	0·2	0·5	0·7	1·0	1·2	1·5	1·7	2·0
0·40	0·2	0·3	0·5	0·6	0·8	1·0	1·1	1·3
0·60	0·0	0·1	0·1	0·2	0·2	0·3	0·3	0·4
0·80	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
1·00	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
1·50	−0·0	−0·1	−0·1	−0·1	−0·1	−0·2	−0·2	−0·2
2·00	−0·1	−0·2	−0·3	−0·4	−0·5	−0·6	−0·7	−0·7
3·00	−0·2	−0·4	−0·6	−0·9	−1·1	−1·3	−1·5	−1·7
4·00	−0·3	−0·6	−0·9	−1·2	−1·5	−1·8	−2·1	−2·3
6·00	−0·4	−0·7	−1·1	−1·5	−1·8	−2·2	−2·5	−2·9
8·00	−0·4	−0·7	−1·1	−1·5	−1·9	−2·2	−2·6	−3·0
10·0	−0·3	−0·7	−1·0	−1·4	−1·7	−2·1	−2·4	−2·8
15·0	−0·3	−0·5	−0·8	−1·0	−1·3	−1·5	−1·8	−2·0
20·0	−0·1	−0·2	−0·4	−0·5	−0·6	−0·7	−0·9	−1·0
30·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·9
40·0	0·3	0·6	0·9	1·2	1·5	1·8	2·2	2·5
50·0	0·5	0·9	1·4	1·9	2·4	2·9	3·4	3·9

of 100 α in the Table suggest that the use of formulae (2) and (3) is in many cases adequate.

George P Cosmetatos
(*March 1976*)

Department of Management Science
Imperial College of Science and Technology
Exhibition Road
London SW 7 2BX

REFERENCES

1. EILON S (1960) Inventory control: a problem in stocking of perishable goods. *Production Engineer* **39**, 210.
2. EILON S (1962) *Industrial Engineering Tables*. Van Nostrand, London.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

Paired Comparisons in Measuring Expected Benefit

THERE are many allocation problems in the public sector which may be formulated as linear programmes, where the aim is to maximise value derived by customers subject to various resource constraints. In a recent study [2], a social services problem (which of service types i should be allocated to potential recipients j) was formulated as

$$\max f = \sum_{ij} c_{ij} \cdot h_j \cdot x_{ij}$$

subject to

$$\sum_{ij} g_i \cdot x_{ij} \cdot h_j \leqslant B$$

and

$$\sum_{i} x_{ij} \leqslant 1 \text{ for all } j.$$

h_j is the number of persons of type j, g_i is the unit cost of service type i, B is the available budget, and c_{ij} is the relative expected benefit derived by person j through allocation of service i.

The c_{ij} values were obtained by asking questions to social workers in a way based on Ackoff's paired comparisons method of utility measurement [1]. Briefly, this involves ranking of all c_{ij} values (recipient type/service type combinations) and subsequent translation to a ratio scale by asking preference or indifference between pairs of packages, including multiple allocations—such as x_{ij} *cf.* $2x_{kl}$. The task of ranking within rows (all services for one recipient type) and columns (all recipients for one service type) is relatively easy, so it is advisable to do this first and obtain as much information indirectly (by transitivity assumptions) before seeking rankings involving elements in different rows and columns. In rare cases, this will produce a complete ranking as in Fig. 1. below.

The matrix gives the row and column ranking of any x_{ij}. For any element, the number in brackets indicates its row rank while the other number provides the column rank. Thus, in this instance, a complete ranking is possible. Where this is not so the second stage—asking paired comparisons like x_{ij} *cf.* x_{kl}—must be carried out.

Since most matrices will be roughly square, we wish to determine, for an $n \times n$ matrix, the maximum possible number of pairs whose relative rankings could be unknown (max U) after row by row and column by column ranking. It is assumed that there are no loops in the ranking matrix, since if a judge produces a loop he can be asked to resolve it.

If elements of each row are arranged in order of decreasing size from left to right, any feasible set of columns comprises n lines, each passing through one element from each row. Now, it is intuitively reasonable that arrangements with many start points and end points, and with many elements being ranked the same in their columns as in their rows should provide large values of U. Thus the following arrangement, where there are n start and end points, and every element is ranked the same in its row as in its column, may be close to max U. For this arrangement, the value of U can easily be computed by summing of series, and is

	Service type A	B	C	
Recip. type	1(1)	3(2)	1(3)	(1)
	2(1)	2(2)	2(3)	(2)
	3(1)	1(2)	3(3)	(3)

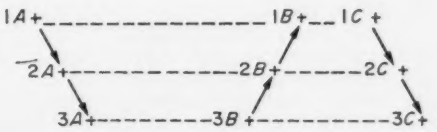


FIG. 1.

Vol. 4
1976

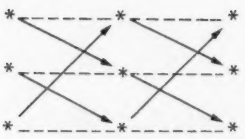


FIG. 2. ($n = 3$).

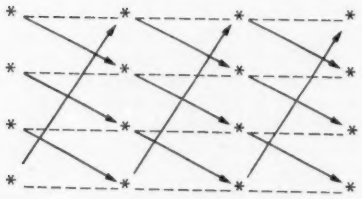


FIG. 3. ($n = 4$).

$$U = \frac{n^2 \cdot (n-1) \cdot (2n-1)}{6}.$$

It can be demonstrated that this is max U, as follows. For an $n \times n$ matrix, there are $\frac{1}{2}.(n^2).(n^2-1)$ pairs of elements. Ranking by rows and columns separately provides information on $(n^2).(n-1)$ paires—not counting any which may be deduced by transitivity. The minimum number of pairs which can be ranked through transitivity is

$$\begin{aligned} &n \cdot [1 + 2 + 3 + \ldots (n-2)] \\ &+ n \cdot [1 + 2 + 3 + \ldots (n-3)] \\ &+ \ldots n \cdot [(n - (n-1)] \end{aligned}$$

since every element must be connected at least once to every other row. It is these column connections between elements in different rows which produce the rankings by transitivity.

The above function sums to $\frac{1}{6}.(n^2).(n-1).(n-2)$, so that the maximum number of unranked pairs is

$$\begin{aligned} &\tfrac{1}{2} \cdot (n^2) \cdot (n^2 - 1) - (n^2) \cdot (n-1) \\ &- \tfrac{1}{6} \cdot (n^2) \cdot (n-1) \cdot (n-2) \end{aligned}$$

which is $\frac{1}{6}.(n^2) \cdot (n-1) \cdot (2n-1)$ as obtained for the trial solution above.

There are other patterns of links that produce max U, but they may all be derived from the above arrangement by changing the order of the rows or columns. It can be proved that only such row or column changes will produce max U; an essential feature of the set of arrangements being that each element has the same rank in its row as in its column. It may also be shown that when given a row and column of the matrix, there is a unique assignment of ranks to produce the max U solution.

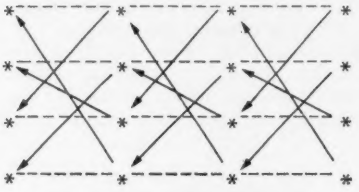


FIG. 4. *A max* U *arrangement.*

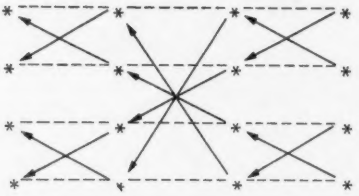


FIG. 5. *A non-max arrangement which cannot be derived from the arrangements opposite.*

REFERENCES

1. ACKOFF RL and CHURCHMAN CW (1954) An approximate measure of value. *Ops Res.* **2**, 172.
2. HINDLE D (1975) Local authority social services—the allocation problem. Unpublished Doctoral Thesis, University of Lancaster.

D Hindle[1]
R Sergeant[2]

Greengarth Hall
Holmrook
Seclafield
Cumbria CA19 1UL

(March 1976)

[1] Now at Federal University of Santa Maria, Brazil.

[2] Now with British Nuclear Fuels Limited, Sellafield, Cumbria UK.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

How Many-Sided are Shadow Prices at Degenerate Primal Optima?

STRUM [6] has given a standard-maximum-LP-problem in production planning, in which the primal optimum is degenerate. In Strum's example one basic variable is zero; two primal and dual feasible and hence optimal tableaux with different dual values are considered. The solution of two slightly modified LPs shows that the value of having one more unit of time on machine 3 is 0·5, while the value of having one unit less is 1·4. Therefore the shadow price is said to be 'two-sided'. Related ideas can be found throughout the literature.

Eilon and Flavell [2] have shown that there exists an additional basis for the primal optimum in Strum's example; because of the infeasibility of an allied dual value the authors ascertain a violation of the Necessity Theorem for defining an optimal point in a LP. An interesting correspondence between dual infeasibility and the effects of a constraint's omission is established. Moreover, 'Two-sided Shadow Prices' are rejected as an oversimplification, because any marginal variable at any degenerate vertex could take a number of different values and therefore the shadow prices were 'many-sided'.

Considering the economic content of dual variables in the nondegenerate case, we can distinguish at least two different kinds of interpretations. On the one hand dual variables are used to evaluate scarce resources, to use these values in decentralised decision making (e.g. see [5, pp. 345–350]) and to persuade the executive staff of the optimality of the central planned optimum ([4, p. 141]). In this sense *any* tableau that leads to the primal optimum can be used to establish 'prices'. In doing so one could choose even a tableau that is dual infeasible like Eilon–Flavell's third solution, but negative prices for scarce resources might result in difficulties in the understanding of the allocation mechanism. On the other hand dual values enable the valuation of actions not taken into account in the original problem. We shall consider the effects of capacity variation as is done most frequently and especially in the papers of Strum and Eilon–Flavell. Employing the usual formulas of sensitivity analysis (see e.g. [3, p. 146]) one obtains a closed interval $[b_t^-, b_t^+]$ in which the available units of the resource must lie for maintenance of primal feasibility of the base and therefore for significance of the dual value of the associated slack variable. At a nondegenerate optimum we have for the units b_t originally available from resource t

(a) $b_t^- < b_t < b_t^+$.

At a degenerate optimum three other cases might occur:

(b) $b_t^- = b_t < b_t^+$,

(c) $b_t^- < b_t = b_t^+$,

(d) $b_t^- = b_t = b_t^+$.

Which of the four cases results depends on the sign of the coefficients $\bar{a}_{tp}$ of the associated slack variable p in an optimal basis $\bar{\mathbf{B}}$. Let $M \equiv \{i | \bar{b}_i = 0\}$ be the set of indices of all 'degenerate rows'. The dual value $\bar{c}_p$

(a) can be interpreted as at a nondegenerate optimum, if $\bar{a}_{kp} = 0$ for all $k \in M$;

(b) gives the *increase* of the objective function if b_t were expanded to $b_t + 1 \leq b_t^+$, if

$$\bar{a}_{kp} \geq 0 \text{ for all } k \in M \text{ and } \bar{a}_{rp} > 0 \text{ for at least one } r \in M; \quad (1)$$

(c) gives the *decrease* of the objective function if b_t were reduced to $b_t - 1 \geq b_t^-$, if

$$\bar{a}_{kp} \leq 0 \text{ for all } k \in M \text{ and } \bar{a}_{rp} < 0 \text{ for at least one } r \in M; \quad (2)$$

(d) has no significant economic interpretation, if there exist $r \in M$ and $s \in M$ such that

$$\bar{a}_{rp} \cdot \bar{a}_{sp} < 0. \quad (3)$$

Starting from a primal and dual feasible solution, one can find another feasible and

optimal basis if a dual simplex iteration with pivot element $\bar{a}_{kp}$, $1 \in M$, is performed ([7, p. 84]). If the primal optimum is unique (i.e. no *alternate* solution exists), each fully employed resource will have a positive dual value $\bar{c}_p > 0$ if the corresponding coefficients $\bar{a}_{kp}$ fulfil (2). If a slack variable can be made basic by one or more dual simplex iterations its coefficients fulfil (1) and therefore an expansion of the accompanied resource will not improve the value of the objective function. If a slack variable p cannot be made basic by dual simplex iterations a solution fulfilling (1) and $\bar{c}_p > 0$ exists.

If, in a basis $\bar{B}$, condition (2) or (3) holds for column p and a solution that satisfies (1) is wanted, one can perform a dual simplex iteration on row $1 \in \bar{N} = \{i | \bar{a}_{ip} < 0$ and $b_i = 0\} \subset M$. Carrying out this iteration with pivot $\bar{a}_{iq} < 0$ from base $\bar{B}$ to $\tilde{B}$, a nondegenerate dual guarantees $\tilde{c}_q > 0$ and therefore $\tilde{c}_p < \bar{c}_p$. If $\bar{N}$ is nonempty a further iteration must be performed. Finally a basis $\tilde{B}$ will be attained in which $\tilde{a}_{kp} \geq 0$ for all $k \in M$ holds. The shadow price of a resource for expansion will thus never exceed the price of its reduction. (For a formal proof see [1]). In a similar way dual simplex iterations can be performed to attain a base $\tilde{B}$ with $\tilde{a}_{kp} \leq 0$ for all $k \in M$.

Note that to obtain the information given by the nondegenerate optimal tableau one has to judge several degenerate tableaus. Even if a large number of different dual values is associated with a resource no more than two of them are meaningful in the evaluation of capacity modifications. Therefore from an economic point of view one should not replace the notion of two-sided shadow prices by that of many-sided prices.

REFERENCES

1. Balinski ML and Baumol WJ (1968) The dual in nonlinear programming and its economic interpretation. *Rev. Econ. Studies* **35**, 237–256.
2. Eilon A and Flavell R (1974) Note on "Many-sided shadow prices". *Omega* **2**(6), 821–823.
3. Gass SI (1969) *Linear Programming*, 3rd ed. McGraw Hill, New York.
4. Hax H (1974) *Entscheidungsmodelle*. Rowohlt, Reinbek bei Hamburg.
5. Kwang C and Wu Y (1971) *Mathematical Programming and Economic Analysis of the Firm*. Intext Educational Publ., Scranton.
6. Strum JE (1969) Note on "Two-sided shadow prices." *J. Acc. Res.* **7**, 160–162.
7. Varga J (1974) *Praktische Optimierung*. Oldenbourg, München.

Dr G Knolmayer
(*May 1976*)

Institut für Betriebswirtschaftslehre
der Universität Wien
Türkenstraße 23/11
A-1090 Wien
Austria

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 4, 1976. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

AßMANN K (1971–72) *Zustand und Entwicklung des Städtischen Handwerks in der Ersten Hälfte de* 19 *Jahrhunderts*. PhD Thesis, Georg-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 2400 Göttingen, Prinzenstrasse 1, Germany)

CLEARY MJ (1971) The Application of a Linear Programming Transportation Model in Determining Comparative Location Advantages. Doctoral Thesis, University of Nebraska. (Available at the Love Library, University of Nebraska, Lincoln, NE 68508, USA)

GAS B (1971–72) *Wirtschaftlichkeitsrechnung bei Immateriellen Investitionen*. PhD Thesis, George-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 2400 Göttingen, Prinzenstrasse 1, Germany)

GROßHEIM G (1971–72) *Die Auswirkungen des Zentralamerikanischen Gemeinsamen Marktes auf die Industrien der Teilnehmerstaaten*. PhD Thesis, Georg-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 2400 Göttingen, Prizenstrasse 1, Germany)

GURPERSAUD (1974) Factors Affecting the Establishment of Viable Economic Farms in the Wauna Area of Guyana. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

HAEBERLIN F (1972) *Die Eingliederung Junger Flüchtlinge als Problem der Spätsozialisation. Eine Empirishe Untersuchung*. PhD Thesis, Georg-August-Universität Göttingen. (Available at: Neidersächsischen, Staats- und Universitäts-Bibliothek, 2400 Göttingen, Prizenstrasse 1, Germany)

HANSEN P (1971–72) *Die HandelsgerichteteAabsatzpolitik der Hersteller im Wettbewerb um den Regalplatz—Eine Aktionsanalytische Untersuchung*. PhD Thesis, George-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 2400 Göttingen, Prizenstrasse 1, Germany)

HEDLEY HA (1974) An Analysis of the Spatial Organization of the Ontario Feed Manufacturing Industry. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

HOWE KM (1974) The Impact of the Economic and Regulatory Setting on the Public Utility's Cost of Capital. Doctoral Thesis, University of Nebraska. (Available at the Love Library, University of Nebraska, Lincoln, NE 68508, USA)

KNOD EM (1974) The Development and Testing of a Model of Employee Turnover. Doctoral Thesis, University of Nebraska. (Available at the Love Library, University of Nebraska, Lincoln, NE 68508, USA)

MOHAMMED G (1972) An Econometric Analysis of the Supply of and Demand for Sugar in Canada. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

Vol. 4
1976

PAGO VK (1973) Simulation as a Farm Management Planning Tool. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

RICHARDSON GC (1972) An Analysis of the Demand and Price Structure of the Market for Maritime Potatoes. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

SMITH DB (1971) The Economics of Inter-Energy Competition in the United States: An Empirical Investigation. Doctoral Thesis, University of Nebraska. (Available at the Love Library, University of Nebraska, Lincoln, NE 68508, USA)

STONE PM (1973) A Simulation Approach to Optimal Resource Allocation for White Veal Production in Ontario. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

TARTE F (1974) A Model of the Feed Purchase Deicision Process. MSc Thesis, Ontario Agricultural College. (Available at the School of Agricultural Economics and Extension Education, Ontario Agricultural College, University of Guelph N1G 2W1, Canada)

VONDERACH G (1971–72) *Industriearbeit und Wirtschaftsgesellschaft in der Bewußtseinsbildung von Studenten.* PhD Thesis, Georg-August-Universität Göttingen. (Available at: Niedersächsischen Staats- und Universitäts-Bibliothek, 2400 Göttingen, Prinzenstrasse 1, Germany)

The following are obtainable at the Yvette Jacobson Library, Management Centre, University of Bradford, Emm Lane, Bradford, Yorkshire BD9 4JL, UK:

HOOLEY P (1973) The New Issues Market: An Examination of the Methods of Marketing a New Equity Issue. MSc Thesis, University of Bradford.

HUTCHINS T (1973) Analysis of Specific Physical Distribution Problems in a Medium Sized Textile Company. MSc Thesis, University of Bradford.

JACKSON C (1972) Cross-Frontier Mergers in Europe. MBA Thesis, University of Bradford.

JENKINS JR (1970) Polemic on the Theory of Buyer Behaviour. MSc Thesis, University of Bradford.

JENKINS PH (1972) The Exit Interview: Methodology and Validity. MSc Thesis, University of Bradford.

JOELS CP (1971) Cost Control and the Speculative House Builder. MSc Thesis, University of Bradford.

JONES AJ (1971) Personnel Management in Small Firms: A Comparison. MSc Thesis, University of Bradford.

JONES CJ (1973) Financial Statements and the Investment Decision. MSc Thesis, University of Bradford.

JONES RS (1972) The Role of Direct Distribution in a Developing Economy. MSc Thesis, University of Bradford.

KARP JB (1972) International Advertising Management. MSc Thesis, University of Bradford.

KELLY JWE (1973) Japanese Impact on the Motor Cycle Market. MSc Thesis, University of Bradford.

KENDALL TR (1970) The Application of Sales Forecasting in the Office Machinery Industry and its Relative Poisition in Organising Productive Capacity and in the Formulation of Long Range Corporate Planning. MSc Thesis, University of Bradford.

KENNEDY SH (1970) The Rationality of the Industrial Buyer—A Study in the Transport Industry. MSc Thesis, University of Bradford.

KEPINSKI JA (1973) Economic Aspect of Sulphuric Acid Production in the United Kingdom. MSc Thesis, University of Bradford.

KIRKBY D (1973) Debt Factoring in the United Kingdom. MSc Thesis, University of Bradford.

LANGDELL JA (1972) Consumer Characteristics in the Man's Outerwear Market. MSc Thesis, University of Bradford.

LEACH JD (1972) Marketing and Information Systems: A Review and Evaluation. MSc Thesis, University of Bradford.

LONGBOTTOM J (1972) A Study of Work Planning in the Electricity Supply Industry. MSc Thesis, University of Bradford.

LUMLEY R (1971) White-Collar Unionism in Britain: A Survey of the Present Position. MSc Thesis, University of Bradford.

MCCANN TW (1970) Evaluating and Improving a Survey Method of Forecasting the Stock Requirements for Fashion Goods. MSc Thesis, University of Bradford.

MCDONALD MHB (1973) The Application of Marketing Theory in a Recovery Situation. MSc Thesis, University of Bradford.

MICO PR (1973) Brand Images. MSc Thesis, University of Bradford.

MILLER KEJ (1971) Stock Management in a Large Light-Engineering Company. MSc Thesis, University of Bradford.

MORTON WJ (1972) A Study of Promotions Within the Construction Industry. MSc Thesis, University of Bradford.

POLLARD R (1973) A Precis of Some Conceptual Problems of Income Determination and Stabilised Accounting, with An Illustration of Their Effect on the Published Accounts of an Engineering Company. MSc Thesis, University of Bradford.

RIDGEWAY CC (1972) A Critical Evaluation of Role Theory. MSc Thesis, University of Bradford.

ROWE DNE (1973) A Research and Development Project Cost Control System for GEC Marconi Electronics. MSc Thesis, University of Bradford.

SLADE J (1973) Education for Marketing and Advertising. MSc Thesis, University of Bradford.

SUGDEN JP (1974) The Uses and Applications of an Annual Budget: An Empirical Case Study of Budgetary Control. MSc Thesis, University of Bradford.

SYKES EA (1972) Credit and the Independent Retailer. MSc Thesis, University of Bradford.

The following are obtainable at the Library, Cranfield Institute of Technology, Cranfield Bedford, MK43 0AL, UK:

GAINES G (1972) Career Development at the International Monetary Fund. MBA Thesis, Cranfield Institute of Technology.

GALLOWAY IW (1972) The Impact of Exhaust Emission Legislation on US Market for Diesel Engined Trucks. MBA Thesis, Cranfield Institute of Technology.

GILLIVER (1970) Scheduling of Jobs on Machines with a Technological Constraint on Routing. MBA Thesis, Cranfield Institute of Technology.

GORLEY PJ (1971) An Investigation into the Operation of the Youth Employment Service in Oxfordshire. MBA Thesis, Cranfield Institute of Technology.

HAMILTON AC (1970) Optimum Water Policy for Fire Fighting. MBA Thesis, Cranfield Institute of Technology.

HARRISON GM (1971) A Survey of the British Educational Technology Market. MBA Thesis, Cranfield Institute of Technology.

HUNT JPV (1970) The Current Position and Recommendation for Future Policy in a Small Furniture Company. MBA Thesis, Cranfield Institute of Technology.

JENKINS JH (1970) A Network Problem with Fixed Costs. MBA Thesis, Cranfield Institute of Technology.

KARVOUNOPOULOS GE (1970) Sampling Techniques for Management Control. MBA Thesis, Cranfield Institute of Technology.

Vol. 4
1976

KAY HR (1970) The Impact of Tape Records on the UK Recorded Music Market, 1970–1980. MBA Thesis, Cranfield Institute of Technology.

MCNEIL A (1970) Constructing a Financial Planning Model. MBA Thesis, Cranfield Institute of Technology.

MITCHELL RG (1971) The Market for Tonic Wine in the United Kingdom—An Assessment of Opportunities for Vinbofer. MBA Thesis, Cranfield Institute of Technology.

MORGAN GJ (1972) An Investment Appraisal of Selected UK Quoted Companies in the Battery Market. MBA Thesis, Cranfield Institute of Technology.

NELBERG RA (1972) Opportunities in the Frozen Food Market in Scotland. MBA Thesis, Cranfield Institute of Technology.

NEVILLE-O'BRIEN JV (1971) A Study of the Scheme of Direct Payments Towards the Provision by General Practitioners of Surgery Premises. MBA Thesis, Cranfield Institute of Technology

NIFAKOS T (1971) Determination of the Optimal Range of Products in a Consumer Industry. MBA Thesis, Cranfield Institute of Technology.

PAPATHANASSIOU M (1971) The Development of a System to Control the Introduction of New Products in a Toy Company. MBA Thesis, Cranfield Institute of Technology.

PARKINSON CR (1970) To Compare the Value of Chemical Production and Chemical Export Performance of UK and Germany. MBA Thesis, Cranfield Institute of Technology.

PATRICK EE (1972) Hardware Distribution in the UK. MBA Thesis, Cranfield Institute of Technology.

PEACE H (1972) Commercial and Industrial Property. MBA Thesis, Cranfield Institute of Technology.

RICHMOND R (1972) Problems of Decentralisation in an International Oil Company. MBA Thesis, Cranfield Institute of Technology.

RIGOPOULOS TA (1971) A Survey of Operational Research in the Hospital Services. MBA Thesis, Cranfield Institute of Technology.

SAUNDERS JA (1971) An Analysis of the World Market for Airport Crash Tenders and the Relationship Between Gloster Saro Ltd. and the Market. MBA Thesis, Cranfield Institute of Technology.

SELBY MJP (1972) A Financial Model of the UK Motor Vehicle Industry. MBA Thesis, Cranfield Institute of Technology.

SHEAD JC (1972) A Survey of the UK Hobby Market. MBA Thesis, Cranfield Institute of Technology.

SWARBRICK JB (1970) Manpower Planning—The Development of a Systematic Approach to the Problem within the Wall Paper Manufacturers Limited. MBA Thesis, Cranfield Institute of Technology.

THOMPSON VOS (1970) A Critical Analysis of the Management and the Environment of a Tannery. MBA Thesis, Cranfield Institute of Technology.

VARLEY A (1970) Report on Marketing Policies of BTR Polymer Fabrication Group Technical Section. MBA Thesis, Cranfield Institute of Technology.

The following are obtainable at the Lyon Playfair Library, Imperial College of Science and Technology, Exhibition Road, London SW7 2BX, UK:

CHEONG HYC (1972) A Marketing Simulation Model of a Mail Order Company, MSc Thesis, Imperial College, London.

DAMTAS CE (1972) Private Company Financing. MSc Thesis, Imperial College, London.

DIAMANTIDES AS (1972) Liquefied Natural Gas, Production Technology, Economics and Future Trade. MSc Thesis, Imperial College, London.

DUNCAN PJ (1971) Productivity Agreements and the Development of the ICI Weekly Staff Agreement. MSc Thesis, Imperial College, London.

EARLEY RS (1971) Ore Purchasing. MSc Thesis, Imperial College, London.

EYRE TJ (1971) A Study of Computing Economics. MSc Thesis, Imperial College, London.

FEILER DL (1971) The Design of a Decision Table Preprocessor. MSc Thesis, Imperial College, London.

FELDLE R (1971) On a Mathematical and Numerical Description of Times-Series. MSc Thesis, Imperial College, London.

FISCHER PM (1971) Impulse Buying. MSc Thesis, Imperial College, London.

FULLERTON-BATTEN W (1971) A Study of the Money Market. MSc Thesis, Imperial College, London.

GOLDSBY DJ (1971) Installation of a Financial Corporate Planning Model using Linear Programming. MSc Thesis, Imperial College, London.

HARRIS RG (1972) A Time Series Analysis of the Gilt-Edge Market. MSc Thesis, Imperial College, London.

HENNEY GBF (1972) Economics Analysis of Plastic Milk Bottles. MSc Thesis, Imperial College, London.

HONEY L (1971) A Study into the Effectiveness of the Pharmaceutical Promotion. MSc Thesis, Imperial College, London.

HUNT RE (1971) A Review of Manpower Planning Models. MSc Thesis, Imperial College, London

KAFANDARIS S (1971) The Diffusion of Operational Research Activities in the UK. MSc Thesis, Imperial College, London.

KLINGHOFFER O (1971) Medical Instruments: A Market Survey. MSc Thesis, Imperial College, London.

LANE DC (1971) An Investigation into the Introduction and Implementation of a Productivity Agreement at Seventeen Chemical Works. MSc Thesis, Imperial College, London.

LEE JTH (1971) A Probabilistic Activity Model for Network Analysis. MSc Thesis, Imperial College, London.

LIVESEY IA (1971) Student Selection and Information Retrieval System. MSc Thesis, Imperial College, London.

MCKEAN JM (1971) The Decision Making Process in Research and Development. MSc Thesis, Imperial College, London.

NAGARAJAN R (1971) Aspects of an Implementation of a Corporate Financial Planning Model. MSc Thesis, Imperial College, London.

PATEL LC (1971) Problems in Share Valuation. MSc Thesis, Imperial College, London.

PATIENT DIT (1971) An Algorithm for Advertising Media Scheduling. Imperial College, London.

PATRICK AW (1971) Some Aspects of Trade Credit. MSc Thesis, Imperial College, London.

PATTERSON RM (1971) An Earnings Evaluation Model for Open-Pit Mining Operations. MSc Thesis, Imperial College, London.

PEDDER AP (1971) A Study of Absence Behaviour in the Mining Industry. MSc Thesis, Imperial College, London.

PERKS RF (1971) A Report on the Gilt Edged Yield Curve. MSc Thesis, Imperial College.

RAYNER AR (1971) Corporate Objectives: A Survey and Critique. MSc Thesis, Imperial College, London.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Vol. 4
1976

EDITORIAL

Planning and Freedom

CONVENTIONAL wisdom tells us that there is a dichotomy between a planning function and the concept of freedom. As economists, sociologists and political scientists point out, the very act of planning implies that men, machines and money are directed towards the achievement of certain prescribed ends and this inevitably means that freedom of action is denied to those who wish to pursue other ends. The ultimate realization of central direction is epitomized in a planned economy, where the allocation of resources to the major sectors of the economy is centrally administered, with output and various performance measures set for each sector, as well as the materials flow from sector to sector. Instead of each industry having the freedom to determine its own objectives, its level of output and the range of its products, the central plan determines the objectives and targets that should be attained; instead of allowing the customer to exercise a choice between a wide range of goods, the central administration determines the range of choice, if any, that should be provided; instead of the price mechanism adjusting the balance between supply and demand, the central authority decides on the shape and form of final demand. All these actions of centralized planning are interpreted, and rightly so, as measures bound to restrict freedom of choice: freedom of customers to decide what to buy, freedom of manufacturers to decide what and how much to produce, freedom of labour and finance to move from one industrial sector to another, freedom of management to introduce technological change with its implications for capital expenditure and manning of facilities.

The argument of those who favour a planned economy is not that loss of freedom does not occur, but that the loss is greatly outweighed by the benefits to society as a whole from coordinated economic activities. The ability of each industrial sector to proceed relentlessly with production according to a given plan, immune from the vagaries of the market place and unpredictable actions of competitors, is thought to ensure that scarce resources are efficiently utilized and that long term payoffs must then accrue. It is for these reasons that advocates of a planned economy are prepared to acquiesce with what they regard as a marginal loss of personal freedom for the good of all.

On the other hand, those who believe in a free economy see the need to preserve freedom as a paramount objective. First they argue that a planned

economy is unable to allocate resources in a sensible way, that a centralized bureaucracy inevitably becomes cumbersome and totally incapable of controlling effectively a highly complex and intricate system such as the national economy. But apart from these largely pragmatic arguments, they postulate that what suits the individual is what matters in society, and that the sum total of individual choices and decisions—influenced as they must be by price and competitions—represents the best outcome for an economic system whose resources eventually get directed to meet the needs and aspirations of those individuals. That such choices take time to permeate through the system and that they can lead to many anomalies, such as the over-provision of productive capacity in some industries, is accepted as an inevitable outcome, but no bigger price to pay—so it is argued—than in a planned economy, where serious time lags are just as much in evidence and where matching capacity to demand does not always work out as smoothly as one would expect. The fundamental difference between the free-for-all advocates and the planners is that the former believe in the whim of the consumer while the latter prefer to rely on the wisdom of the planner, the former see optimization as a process that moves from the bottom upwards (what is good for the individual is good for society), the latter insist that optimal allocation of resources can only make sense at the macro level (what is good for society is good for the individual), the former are more sensitive to short term changes in the market place while the latter tend to ignore them as irrelevant to a long term vision of how society should work.

For the sake of simplicity, the two philosophies are presented here as extremes, although their pure form tends to become diluted with the passage of time. More and more planned economies divert resources to consumer goods and provide some choice, not only between specific goods but between different categories of personal expenditure, so that the shape of final demand becomes less pre-determined, and with it pressures mount to reallocate resources to accommodate this demand. Similarly, more and more free economies resort to central control mechanisms first with legislation intended to keep the market place free and to protect the consumer (such as anti-monopolies and anti-dumping laws, and regulatory processes to govern quality, manufacture and distribution of goods), then with fiscal measures as well as prices and incomes controls. But in spite of this dilution of the two extremes, they remain wide apart, both in terms of their basic philosophies as in their attitudes to the planning function.

Now, the two philosophies alluded to are very well known as representing the opposing political beliefs of their adherents, and it would not have been necessary to summarize them here had it not been for a recently published paper by van Gigch entitled 'Planning for Freedom' [1]. As the title suggests, the author argues that "freedom can be planned", that a planning system can enhance as well as hinder the exercise of freedom, and he proceeds to show that different types of planning systems (he compares four: the liberal model of democracy,

the conservative model of democracy, the broker rule model, and the Marxian socialist–communist model) can each claim not only to be 'democratic' but also to provide certain kinds of freedom, so that the problem of comparing the amounts of freedom thereby accorded remains unresolved. He concludes that "there is no inherent dilemma between planning and freedom" since "more planning does not necessarily lead to less freedom".

The first fundamental question that needs to be asked is what is meant by 'freedom' and by 'planning'. Van Gigch dwells at some length on the first and points out that some writers distinguish between 'negative freedom', which is based on the notion that only the presence of something renders a person unfree, and 'positive freedom', which suggests it is the absence of something which makes him unfree, although van Gigch appears to accept the view that this distinction is not particularly helpful. He then proceeds to argue that freedom always involves *assumptions* (by planners), *conditions* of the system (including the role of rationality in planning, equality and participation of individuals, the prevailing technology, and attitudes to change), and *opportunities/barriers*, and suggests that different combinations of the three ingredients in this triad can explain the varieties of freedom encountered in their respective political systems.

Vol. 4
1976

I am less inclined to dismiss the concepts of 'negative freedom' and 'positive freedom' as inappropriate for our discussion. The terms as such are perhaps not that attractive, but they highlight the proposition that freedom essentially involves the *provision of choice* and the *exercise of choice*. The provision of choice implies that an individual is given a range of alternatives, the exercise of choice means that he is able to make a selection without threat or coercion. The two elements are obviously interconnected: if a person is given a wide spectrum of alternatives but is forced to choose only one, the outcome for him is the same as when he is assured that he has freedom of choice but is offered no alternatives; it is, nevertheless, worth stating the two conditions, particularly as the attitudes of the individual in the two situations just described may well be rather different.

It is important to point out that the provision of freedom to one individual or group of people may result in a denial of freedom to another. For example, if pedestrians are allowed to walk freely on motorways, the freedom of the motorists to drive at any speed is thereby curtailed, similarly equal opportunities for all implies erosion of relative privileges for the few. The measurement of a change in total freedom in a given society following any regulatory measures must therefore take account of gains to some and losses to others, and even take note of the fact that particular individuals can both gain and lose as a result.

Now, *planning* is a manifestation of a regulatory process, it is concerned with the allocation and utilization of resources, it directs the activities of groups and individuals with the object of meeting certain prescribed targets. In what sense can 'planning for freedom' be regarded other than a contradiction in terms?

The answer lies in the proposition that planning culminates in the provision of choice, while the political–social system in which individuals live determines the exercise of choice. To the extent that a planning function does widen the choice for the majority of the members concerned then 'planning for freedom' becomes a legitimate concept.

So far we have been concerned solely with macro-economic planning for a national economy. But what about planning in an industrial enterprise? What fundamental philosophies or attitudes can we discern at that level, and what does planning imply for freedom in that context?

There have been those prepared to extend the argument from the macro to the micro level and to suggest that if planning is regarded as bad for the economy as a whole it must be equally undesirable at the level of the firm, and so we have witnessed over the last two or three decades entrenched opposition from some quarters to the very idea of corporate planning. If you are sufficiently agile to adapt to a changing environment and take care of the short term, so it was argued, the long term will take care of itself. These attitudes towards planning in the firm have undergone a marked change, and many captains of industry no longer see a dichotomy between their view that the national economy should be left to market forces, free from governmental controls and interference, and their support for a strong corporate planning function within their own organizations. This apparent dichotomy cannot be entirely explained by the proposition that at any level in a hierarchy the natural tendency of an individual is to deplore being planned for but to justify planning for his subordinates. A major factor in current attitudes to corporate planning is the mounting pressures to utilize resources effectively (pressures generated by customers, shareholders and competitors) and the realization that some form of direction (such as corporate planning) can usefully replace the constant and irrevocable erosion of autocratic management in a democratic society.

Now, it would be tempting to conclude that in that respect one can talk about 'planning for freedom' within the firm, that planning provides a mechanism for freeing the enterprise and its constituent parts from the shackles of autocratic rule, that any move towards a pluralist managerial system, however tentative, is a realization of greater freedom for individuals within it. But apart from that limited sense (and it is limited if one considers the totality of industry, where old-fashioned autocracy is rapidly receding), there can be little dispute that planning represents a denial of freedom of action, that a corporate plan with its inherent targets and constraints must be accompanied by an increasing degree of prescription for various sections and levels of the enterprise, that centralized planning can and often does lead to inflexible bureaucracy, and that even a pluralist bureaucracy can and does encroach on individual freedom.

This is not to say that opposition to corporate planning is justified. Admittedly, there are many counts on which current corporate planning can be

criticized—such as rigidity, remoteness, ineffectual methodology—but anyone suggesting that a modern industrial enterprise can afford to amble along without a coherent strategy or purpose, is either too naïve or is blind to the increasing level of complexity that industry has to contend with. Corporate planning is here to stay, but—as in the case of governmental regulatory processes—it needs to be performed with wisdom and restraint to provide an appropriate balance between general direction and the attainment of corporate objectives on the one hand and the preservation of some freedom to the constituent parts on the other.

SAMUEL EILON
Chief Editor

REFERENCE

1. VAN GIGCH JP (1976) Planning for freedom. *Mgmt Sci* **22**, 949–961.

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Vol. 4
1976

Feedback

Responsibilities of the Management Scientist

PROFESSOR EILON'S editorial [4] on the Responsibilities of the Management Scientist has a number of themes at several levels. Some points I welcome. Others I would challenge.

Apparent conflict

First, agreement. I welcome the points Eilon makes on conflict of interest in organisations. There has been much argument on this between Ackoff and critics [1, 2, 3, 5]. Ackoff appears to feel that most conflict is apparent conflict and that "Most conflict that appears to be unresolvable at one level of means or needs is subject to resolution at a higher level of desirability" [2, p. 96]. He goes on to advocate that the management scientist should cure this apparent conflict. As Eilon points out, the participants within any organisation, particularly a large industrial one, have differing goals, as do the people the organisation affects, such as its customers and people in the surrounding environment. These individuals themselves have a multitude of goals which may well not be compatible and it is less likely still that the whole collection of goals will be. It is, indeed, difficult to define any goal that all participants would accept.

'Greater efficiency' of the organisation is a goal that is often thought to be unambiguously desirable. This begs the question of how one defines 'efficiency'—profitability, use of least resources?—and also the time scale. In the short run, greater efficiency could mean redundancy for certain employees, or bankruptcy for a competitor. There may be ways by which the management scientist could both get greater efficiency and preserve the jobs of the employees—I would argue that he should try—but these two goals will not always be obtainable simultaneously. To increase the safety of a process to cause no damage to employees' health and at a cheaper cost will be beneficial to the firm and to the employee, but the most common case is where increased safety is bought at a greater cost. In many cases, such as with the dangers of asbestos, risks have been taken at work with employees' lives that would not be taken at home with an individual's own life. One could argue that in the long run the increased cost of safety is to the benefit of the firm but in the face of competitive pressures it may not be so in the short term, whereas in the short or long term, the safety of the employee is of vital interest for him/her.

The scope for ingenuity in finding solutions that will satisfy the different participants and resolve conflicts is greater than many people give credit for—particularly if one can persuade people to adjust the time scales in which they work—but it is still limited. There are real conflicts—as everyone who works in an industrial organisation will know only too well.

Responsibility to the organisation

Next, disagreement. Eilon states that "Some critics launched a vitriolic attack on Ackoff for asserting that 'Parts of an organisation have a responsibility to the whole' " [4, p. 154]. At a presentational level I do not see the criticism as 'vitriolic'. However, at a substantitive level the comments made in this article [3] do not bear the implications that Eilon draws from it of leading to "a recipe for total anarchy". Our statement was, "responsibilities do not arise by mere virtue of membership of a system but only to the extent that the membership is willing, wholehearted and uncoerced". One should expect more responsibility as the choice is a freer one. At one extreme, in the example of the prisoners in a concentration camp, one would hope for total disloyalty and active opposition. In this circumstance, responsibility to the organisation—as has happened in some cases in camps—is a sign of accepting

the horror rather than fighting against it. In the case of black gold miners in South Africa the choice has been a little freer, but with the almost total constraints on alternative employment and the need to live, there is little real choice involved and, as stated, "one would not expect a deep sense of responsibility towards their employers". In a society which offers more various employment opportunities one would expect a greater responsibility. Even then, in societies like the Soviet Union where the state directs employment to a large extent, one would expect little responsibility—and from all accounts, this is what one gets.

In Western societies also, scope for real freedom in choosing between real alternatives is less than many would like to believe. Again, as a result, one gets rather less responsibility than many would like to see. With the present heavy unemployment, it may be a case of any job or none. But apart from this, much of the apparent choice is on which large bureaucratic organisation to join —state, local authority or multinational private company—with only limited freedom on a different type of work. A growing number rebel—'squat' or live on social security, but this is not an option open to many. Despite Eilon's remarks about top management being trapped in their exalted cocoons they do have more freedom than the shop floor worker for alternative types of employment even though they may not wish to exercise this. One would therefore hope for a greater degree of responsibility on their part—this is one reason they are paid more.

The answer to how much responsbility is due to an organisation in return for the real freedom one had in joining it would differ depending on political and social beliefs. However, the concept that increased freedom requires increased responsibility and conversely, decreased freedom brings with it decreased responsibility is one that most would support.

One very important side issue is the question of responsibility to fellow human beings within the organisation. This is always there, no matter what the organisation. One should try to respect and respond to the humanity in others even though one is fighting against them (as in the last war). My complaint is that institutional arrangements often make this very much harder—one is driven to looking at the 'pure economics' and not the effect on the individuals. I would therefore strive for a society in which these constraints are lessened, where people have greater freedom to choose and where they would expect greater responsibility. This is not an easy option. Freedom demands a great deal of the individual.

Professionalism

The management scientist fits within this total picture. By virtue of his practice one would hope that he would have greater insight into the different objectives and different goals within the organisation as well as an ability to suggest alternative ways of attaining them. However, he often wishes to believe that he is doing good for all and thus tries to pick out goals, such as greater efficiency, that are thought to do this. In one sense this is very laudable but it can lead, and in my experience, often does lead to considerable self-deception. It should be the professional duty of the management scientist to see the different interests involved, and to see how his actions are affecting them, even though he may be forced by virtue of his employment to try to satisfy the goals only of some of the participants rather than all of them. It is important that he does not pretend a pseudo objectivity and claims he is serving the interests of all. Political and social views differ amongst management scientists and one would not expect to find them all on one side. One ought to expect an awareness of whose interests they are serving. One would also hope that management scientists as a whole would try to set up some form of help to those who are not able to buy the service of management scientists, and that there would be more support for the suggested 'institute for critical operational research'—or perhaps alternative suggestions to meet this need.

Thank you, Professor Eilon, for raising these questions.

REFERENCES

1. ACKOFF RL (1974) The social responsibility of operational research. *Opl Res. Q.* **25**, 361.
2. ACKOFF RL (1975) A reply to the comments of Keith Chesterton, Robert Goodsman, Jonathan Rosenhead and Colin Thunhurst. *Opl Res. Q.* **26**, 96.
3. CHESTERTON K, GOODSMAN R, ROSENHEAD J and THUNHURST C (1975). A comment on Ackoff's "The social

responsibility of operational research." *Opl Res. Q.* **26**, 91.
4. EILON S (1975) Editorial—responsibility of the management scientist. *Omega* **3**(5), 513–517.
5. ROSENHEAD J (1976) Some further comments on the social responsbility of operational research. *Opl Res. Q.* **27**(1), II 266–272.

BK CHESTERTON

"Firlie"
Chestnut Avenue
Guildford
Surrey
UK

Vol. 4
1976

THE RECENT editorial by Eilon [4] puts forward a number of ideas, some of which appear to be in conflict with those argued in a previous, important and somewhat controversial paper by Ackoff [1]. With most well prepared and ably presented arguments it is not difficult to find oneself in agreement with both sides which could mean that both parties are presenting special cases within a common problem structure. I propose, in this short note, to suggest a way in which this might be demonstrated.

Ackoff and Eilon both start from a system within which there are numerous interacting parts. Those parts of particular concern to the management scientist/operational research worker are the organisation itself, the employees of the organisation (which usually means several distinct but interacting parts again) and society at large. Discussion now centres on the responsibility that the management scientist has for the effects that recommendations that he may make have on these different parts. In the past there has been a tendency to identify with just one of these parts—usually the organisation itself—which leads to neat and mathematically satisfying optimising techniques. I think that we now recognise that this is not enough, although it is possible to rationalise this revised way of thinking in a number of ways, one at least which brings the argument full circle to sound rather like Charles E. Wilson's "What's good for the country is good for General Motors and what's good for General Motors is good for the country". I do not propose to go further into this change in attitudes, but will discuss the specific point at which Ackoff and Eilon differ fundamentally.

When a decision has to be taken which will affect parts of the system with different and possibly conflicting goals, Ackoff's approach is to discover ways in which actions can be taken serving the interests of all parts. The approach, therefore, assumes the existence of what I will call a 'universal benefit set', that is a set of outcomes where all concerned parts are better served. It is the task of the management scientist to exercise his ingenuity to find an action which will lead to a member of this set. Ackoff quotes a number of examples where he has done just that, and moreover claims that he has never failed to find such an action. There are, however, a number of questions unanswered. For example, the set of outcomes in this universal benefit set must necessarily include outcomes which in the view of at least one of the concerned parts fall short of optimum. Thus a certain amount of reconciliation may still be necessary. More importantly, though, the approach assumes that the management scientist has a clearly defined and accepted view of the real goals of all parts of the system. For example, one of Ackoff's illustrations concerns competition between beverage producers, where one of them is manufacturing a cheaper product containing a syrup with a solvent that may be injurious to health. The action eventually taken brought about legislation which, by means of ingredient identification on the product label, forced the cheaper competitor to discontinue use of syrup. The claim is then made that this served the interests of the original company and of the general public. In a written reply to Chesterton *et al.* [3], Ackoff subsequently [2] argued that there too the interests of the competitor and of the general public were better served by the removal and substitution of the syrup.

Perhaps in this case Ackoff is correct, but we should be aware of the pitfalls in assuming values taken by others. There could be many who would accept the health risk (and it is not a certainty) involved in a cheaper product. Does the management scientist have the right or the knowledge to preclude this? There are a number of cases—motor vehicle safety, alcohol and tobacco consumption, rock-climbing and pot-holing—where people are quite prepared to take risks with their well-being to achieve an aim which most of us would regard as valueless. In the com-

mercial environment of the coal mining industry it could sensibly be argued that management decisions should be aimed to eliminate the dangerous and unhealthy practice of underground work. There is little evidence that this would be acceptable to several of the concerned parts.

My own reaction to Ackoff's claim of always finding a member of this universal benefit set was one of respectful scepticism. The set is defined relative to the position at the time when the decision has to be taken. Once it has been taken and a point within the set realised, what next? No problem is solved once and for all like a theorem in Euclid. To take a further step in the manner that Ackoff proposes, it follows that conditions must have changed, either by external forces, or engineered by a change in attitudes internal to the system (I discount the possibility that an even better decision was possible initially). Failing this change in conditions, some compromise will have to be found. Let me illustrate by extending the Ackoff 'female inspectors' example using what I suggest is a plausible follow-up. The original problem was to improve the quantity and quality of inspection by female workers in a company manufacturing precision-made parts. This was achieved by introducing a scheme of a 'fair day's work' whereby the inspectors finished work as soon as their quota of items had been inspected. As a result the efficiency of the inspection facility improved and the employees benefitted. Suppose now that the employees argue that instead of a fair day's work, why not a fair week's? This would enable them to have a full day away from work if they so chose, and it presumably was rejected initially since the uneven outflow of inspected parts would be incompatible with the work load on the rest of the organisation. A further conflict situation has arisen, which can only be resolved in the Ackoff manner by a change of conditions. This change could arise naturally when the level of discontent among the inspectors results in a fall in their output and efficiency outweighing the cost of the external disruption of work load. Alternatively an external change in conditions could be engineered by, for example, increasing the output level designated as a fair week's work. To say that as a result of such changes all parties are better served merely raises the question—better than what?

If Ackoff is implying that the varying interests of different parties can be resolved without compromise, then we are in danger of neglecting an important stage in our development as management scientists. Eilon, on the other hand, takes the more pragmatic view that where there is a basic conflict it cannot be removed, but only abated. The implication of compromise is clear and it is now a matter of the resolution of multiple objectives, stratified in addition by the parts of the system. This itself is an area which has long been neglected by management science research workers, particularly those working in OR. Perhaps this failure is simply because the problem does not lend itself to unequivocal and absolute arguments, but a recent encouraging sign came in a paper presented by Rivett [6], who suggests that we adapt some of the results of DG Kendall [5] where he shows how meaningful maps can be drawn using only generalised distances between locations. Rivett proposes drawing maps of decision outcomes, where the parallel of proximity in a map is similarity in the attractiveness (to the analyst) of the outcomes. Even based on such a simple 'zero-one' measure, it seems likely that we shall be able to make what in some sense will be consistent decisions on the basis of map locations (outcomes) which will represent admissible compromises, and clearly it is here that we would find the universal benefit set.

The above is an attempt to show how a unified approach might be developed. As far as social responsibility is concerned, it is not so much a matter of developing this trait in the management scientist, but rather that he has a responsibility to take into account the social values of people affected by the decision to be taken.

REFERENCES

1. ACKOFF RL (1974) The social responsibility of operational research. *Opl Res. Q.* **25**, 361–371.
2. ACKOFF RL (1975) A reply to the comments of Keith Chesterton *et al.* *Opl Res. Q.* **26**, 96–99.
3. CHESTERTON K, GOODSMAN R, ROSENHEAD J and THUNHURST C (1975) A comment on Ackoff's "The social responsibility of operational research." *Opl Res. Q.* **26**, 91–95.

4. EILON S (1975) Responsibilities of the management scientist. *Omega* **3**, 513–517.
5. KENDALL DG (1971) Construction of maps from 'odd bits of information'. *Nature* **231**, 158–159.
6. RIVETT BHP (1976) Policy Selection by Structure Mapping. Paper presented to the Second International Research Conference on Operational Research, University of Sussex.

G GREGORY

Department of Management Studies
University of Technology
Loughborough
Leicestershire LE11 3TU
UK

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Causes of the Decline in British Merchant Ship-Building and Marine Engineering

AUSTEN ALBU

(*Received March 1976; in revised form June 1976*)

Britain's share of world ship-building has declined catastrophically during the twentieth century. Successive committees of enquiry have attributed this to a loss of competitiveness deriving both from poor production methods and laggardly innovation. An examination of innovation in ship design and construction and in marine propulsion highlights the declining British contribution in these fields. The conservative attitude adopted at various times by the Royal Navy together with restrictive arrangements in engine design and building seem to have been central in this. Additionally, few graduate engineers have been recruited either into management or into research, development, and design. For the future, a co-ordinated policy on research together with greater specialization in markets should be sought.

Vol. 4
1976

THE POST-WAR decline in British ship-building has been described and documented for the last 20 years. To some extent this decline, measured in proportion of world output, has been a reflection of the inevitable decrease in the British owned proportion of the world mercantile fleet, as shown in Table 1.

TABLE 1.

	GRT owned	Tonnage built
1890	52·3%	81·5%
1900	50·7%	63·0%
1910	45·5%	58·5%
1920	41·5%	54·3%
1930	29·3%	57·2%
1950	21·5%	37·8%
1955	18·3%	27·7%
1960	16·6%	15·9%
1965	13·5%	8·8%
1970	11·4%	5·7%
1974	10·7%	3·6%

(Lloyds Register of Shipping).

In the later years, however, the decline became much more rapid as British shipyards became less and less competitive and even British owners ordered ships abroad. British yards failed to take advantage of the enormous expansion in world tonnage which took place after the war, from 69 million tons in 1939 to 222 million tons by 1970, and were building less tonnage in 1970 than they had been in the nineteen-twenties and thirties. This had been partly due to increasing world shipbuilding capacity; but by 1960 Britain had been overtaken by Japan with a 40% share and by the 1970's by Germany with 20%, Sweden 8% and Holland over 5%. Today there are many new countries developing ship-building capacity; among them Spain (now among the two or three largest in the world), Brazil, Korea and Portugal. The countries of Eastern Europe have not, so far, made much contribution to world markets; but when the USSR naval programme is completed, the Russian, Polish and East German yards will add substantially to world capacity. Much of the expansion which has taken place has been for oil tankers which have formed between 70% and 80% of world output and it is the world surplus of tanker capacity which has caused the present severe slump in the ship-building industry which is forecast to last until the 1980's.

The British industry has recently been subject to a plethora of enquiries; the most important being the Department of Scientific and Industrial Research in Britain (DSIR) Report of 1960, [5] the Patton Report of 1962, [8] the Geddes Report of 1966, [9] and the report by the consultants Booz, Allen and Hamilton of 1972 [2]. All the reports emphasise as the main causes of the decline poor management and inflexible labour; but an examination of the source of innovations in both ship design and propulsion machinery, and even more in production methods, suggests a parallel decline in technical competitiveness. Whereas before the first world war most innovations were of British origin, since the end of the second world war there have been very few. If a willingness to generate and accept innovations is a pre-requisite for maintaining a position in a highly competitive industry, the record of British performance compared with that of other countries could be a major explanation for the drastic decline in the relative position of Britain as a leading ship-building nation. It is not always possible to locate the source of a particular innovation and, in some cases, an invention can be made in one country and first exploited in another (the double helical gear and the all-welded hull are examples); but Table 2 gives some idea of the change that has taken place during the last century and a half. The rapid rate of innovation in the nineteenth century was succeeded by a period of virtually no innovation at all. (The Parsons Turbine and Diesel Engine were late nineteenth century inventions which came into use at the beginning of the twentieth century.) More significant, however, is the almost complete eclipse of the British industry as an innovator since the end of the first half of the century in a period which corresponded with the accelerating fall in its world position.

Vol.
197

TABLE 2. MAIN INNOVATIONS AND THEIR ORIGINS

Period	Innovation	Chief country of origin
Nineteenth century (first half)	Iron hull	Britain
	Steel hull	Britain
	Separation of cargo carriers and passenger liners	Britain
	High pressure steam engine	USA
Nineteenth century (second half)	Oil tanker	Britain
	Use of model tanks	Britain
	Sub-division of hull and cellular bottom	Britain
	Screw propeller	Britain and USA
	Compound engines	Britain
	Water-tube boiler	Britain and France
	Powered steering gear	Britain
Twentieth century (first half)	Steam turbine	Britain
	Geared steam turbine	Britain
	Diesel engine	Germany
	Longitudinal framing (the Isherwood system)	Britain
	All-welded hull	Britain (adopted first in USA and Germany)
	Tilting pad thrust block	Britain and USA
Twentieth century (second half)	Gas turbine	Britain (not yet adopted in merchant ships)
	Very large oil tanker	Japan
	Standard ship built in prefabricated units under cover	USA
	Tenth scale drawings and automatic steel cutting	Germany
	Specialised ships (including gas carriers)	USA
	Electronically controlled and automatic engine room	Japan and Sweden
	Advanced steam data for turbines—use of electrical plant manufacturers' designs	USA
	Very large single shaft diesel engines	Denmark and Switzerland
	Stern oil seals	Germany
	Withdrawable stern gear	Britain—followed by USA and Sweden (none so far fitted by British builders)
	Keyless bore propeller	Britain
	Large ducted propeller	Japan and Norway

Vol. 4
1976

INNOVATIONS IN SHIP DESIGN SINCE THE FIRST WORLD WAR

Between the two world wars there were no radical changes in ship design. It was the period in which the world's tanker fleets grew slowly, prior to the great expansion of motor and air transport and of oil transportation. The Admiralty virtually gave up innovation and the industry seems to have followed. There was, however, at the beginning of this period the start of one significant innovation; and that was the all-welded hull. The first example of a ship so constructed was launched by Camell Laird in 1920 and lost in a collision in 1937 but the cautiousness of the Admiralty and the classification societies delayed its rapid introduction. A British invention, the coated welding electrode, made possible the general adoption of welded connections in the 1930's. The Germans took advantage of the reduction in weight of welded ships to use this method to circumvent the London Naval Treaty of 1932 which limited the weight of ships to be constructed by the signatories. The British industry was the last to continue to use rivetting for the assembly of ships' hulls.

Since World War Two there have been rapid changes following new trends in both design and production methods.

First, the size of ships has grown and they have been built for higher speeds; this has been especially so in the case of oil tankers. Secondly there has been a great increase in specialisation; ships being built, for instance, as Liquid Natural Gas Carriers; and as Product Carriers, Pallet Ships, Container Ships, Barge Carrying Vessels, Roll-on–roll-off Vehicle Transporters and as Ferries.

The concept of sea transport has been undergoing a qualitative change leading to a systems approach in which more attention is paid to the interface between ship and shore. These new ideas are equivalent to the mass production of transport services involving mechanical handling at the ports and technical advances in ship design; the object being to smooth and speed up the flow of goods from factory or plant to consumer.

In introducing these innovations British owners and shipbuilders lagged behind their competitors and, even when they had the advantage of building the first of a new type, failed to maintain this. Although the first ship built to satisfy the extreme conditions required to carry liquified natural gas was a cargo vessel converted at Harland and Wolf's, with American design assistance, only two others have since been built in the United Kingdom. In 1974, 63 Liquified Gas and Chemical carriers were built; 27 in Japan, 9 in Norway, 8 each in the Netherlands and West Germany and none in the UK (as quoted in Lloyds Register of Shipping).

In the construction of very large ships, involving the improved technology required for their construction, the Japanese took the lead. Apart from the greater length and draft of these ships other innovations claimed by the

Vol.
197

Japanese were the bulbous bow (possibly first invented in Britain and of value at higher speeds); the cylindrical bow; and the reduction in thickness of steel plate by the use of high tensile steel leading to reductions in weight and cost [1].

An important and radical innovation has been the introduction of computers in association with the engine room: first to detect faults and later to provide fully automatic control from the bridge. In this the Japanese took the lead in the fifties, probably prematurely, with the *Yamoshu Maru*, a 10,467 ton cargo liner which was equipped with some automatic and remote control features. The French built a tanker with a partly automatically controlled engine room in 1960. The first ship with a fully electronically controlled engine room was the *Emma Maersk*, a 60,000 ton tanker built for Danish owners by Kochums of Sweden in 1964. The British and Commonwealth line, which had become critical of British shipbuilders, set up a working party with two builders to design an automatically controlled engine room and the result was the *Clan McGillvray* also launched in 1964. In 1968 P&O's technical department developed with a Japanese builder the first automatically controlled engine room for steam turbines which was successful. This presented a more difficult problem than the control of Diesel engines to which previous efforts had been confined.

Vol. 4
1976

The use of computers for navigation has made less progress. The first ship with a computer controlled anti-collision system, the Japanese *Senko Maru*, a 138,370 ton tanker, was launched in 1970.

Since it was set up in 1967 the Marine Engineering Centre at Newcastle has concentrated on the design of computer controlled ships systems. In radar type navigational aids there have been substantial British developments based on war-time experience.

INNOVATIONS IN PROPULSION MACHINERY SINCE WORLD WAR II

Turbines built to Parsons designs continued to hold the lead up to the Second World War. Some innovations continued to be made; for instance in 1910 the double helical reducing gear made it possible to use high speed turbines of greater efficiency, but it suffered from early failures and fell into disfavour until the Second World War. The Admiralty were slow to adopt improvements but, before 1939, the United States Naval Engineering Branch, partly for security reasons and partly under the influence of an enterprising and aggressive chief engineer, turned to the manufacturers of electrical plant whose volume of research and development, based on a much higher annual throughput of machinery, was much greater than that of the traditional marine engine builders who lacked research facilities. This change faced strong opposition from the established US shipbuilders, such as Bethlehem, Newport News and the New

York Shipbuilding Corporation who had depended for their engineering on Parsons which was a very conservative organisation. There was also strong opposition from within the US Navy itself whose senior officers feared a too rapid advance in steam temperatures and pressures. With the growth between the wars of large central power plants rapid progress had been made by the electrical manufacturing companies in steam pressures and temperatures leading to higher fuel efficiency; but Parsons generating plant manufacture was a quite separate enterprise from its marine engineering. The inadequacy of Parsons marine engineering continued after the war, and may have contributed to the dramatic failure of the turbines before her maiden voyage of the QE2 in January 1969 [7].

The American Navy eventually turned to the smaller shipyards, using independent naval architects to design their ships and installing high pressure and temperature turbines manufactured by the General Electric Company. These had higher fuel efficiencies and lower weights giving them greater cruising range and were also more reliable. As a result they steamed all round British ships at the end of the war in the Far East. British battleships such as the *King George V* burned nearly 40% more fuel at low speeds and had a worse fuel consumption up to high speeds than American ships such as the *Washington*. The US Naval Attaché in London wrote in 1941, "On the whole we are many years ahead of them (the British) technically. This is true in practically every phase of naval engineering. Many of them look with considerable awe upon our modern machinery installations. I also now understand much better than I did before why your old line (The 'Big Three') insisted on attempting to design as well as manufacture most of the machinery used in the ships they build. This practice is quite universal here in England insofar as boilers, turbines, and reducing gears are concerned. I have visited every shipyard on the Clyde, and everyone has his finger in the design of all three of these important units. In the case of turbines, they all consult with Parsons and he in fact controls the design. Insofar as I can see, he has little competition in this matter and this probably explains the lack of advance which has been made in turbine propulsion here. The Admiralty specifications are very restrictive and they apparently have not kept abreast with the developments made in shore plants. This latter they now realise as having been a serious mistake and I think they will follow the example which you set back home from now on" [3].

The Admiralty exerted pressure on the industry to set up the Parsons and Marine Turbine Research and Development Association in 1944; but although it had a well staffed research department the marine engine builders, most of which were adjuncts of the shipyards, were too small to be able to support it and lacked the level of technical management to be interested in the results of research. Meanwhile some of the best designers employed by its member firms left them to man up the new organisation, the intention being to build only to *Pamatrada* standard designs. In 1946 the Admiralty asked Yarrows to co-

operate with the English Electric Company in a research contract on the effects on machinery of increases in steam temperatures, and from this there was formed the Yarrow Admiralty Research Department (YARD). This was responsible for the design of the successful steam propulsion machinery for frigates and in 1952 of designs for gas turbine propulsion and has since carried out studies of the application of nuclear power to ship propulsion. Its services are also available to private owners. However, naval requirements today do not make much contribution to merchant ship machinery design; naval ships use more expensive fuel and are more concerned with reliability than with running cost. This has, for example, led to a split in the design of boilers as well as of engines. A merchant ship has a normal 'sailing' speed and must be designed for maximum economy at that speed. A naval vessel needs high speeds while in action but, in order to gain operational range, uses a 'cruising' speed at which fuel economy is highest. These two requirements involve great differences in the power required (in the ratio of 1·0:0·1) which makes machinery design difficult. The only likely use of naval R&D in the future is in nuclear ship design and possibly in the development of super-conducting electrical machinery for a ship's drive. Work in this country on the latter at the laboratories of the International Research and Development Company of Newcastle-upon-Tyne is as advanced as any in the world; but it will be some years before its feasibility can be finally judged.

Vol. 4
1976

The original and unpublished version of the DSIR Report of 1960 was very critical of British marine engineering. It pointed out that *Pamatrada* had become merely a licensing organisation and that machinery made to its designs had poorer fuel consumption than other designs. Moreover turbines built in Britain suffered many more breakdowns than those built elsewhere. All British shipbuilders were members of *Pamatrada* and paid a levy on all turbines installed, Parsons design or not. Only two builders had at that time installed turbines to other designs: one of them manufactured by Metropolitan Vickers (afterwards AEI/GEC). During the last 10 years the number of ships built with steam turbine machinery has declined radically by comparison with those in which Diesels have been installed. No turbine set to *Pamatrada* design has gone into service since those of the ill-fated QE2 in 1969 and, in fact, only two sets in all since 1966. The largest number installed in British built ships has been of AEI/GEC design; but an increasing number are being built to the design of the Swedish firm of Sta Laval [5].

The trend in Diesel engine development has been towards larger and larger engines operating on a single shaft. Only one British firm, Doxfords, has tried to design an engine of this scale and it has so far not been a success. Attempts to develop a new marine engine by Hawker–Siddeley also came to nothing. In the 1950's the English Electric Company (Napiers) developed an original but over-complicated design, the Deltic, which was used by the Navy for small ships. Between 1950 and 1959 the proportion of large marine Diesels built to

foreign design increased from 25 to 46% [5], but Doxford's position has been falling rapidly. In the last 10 years the percentage of types of Diesel engines of over 10,000 BHP installed in British ships has been: Sulzer 40%, Burmeister Wain 36% and Doxford 23% (from Lloyds Register of Shipping). Moreover there seems little evidence of British engine builders making significant innovations to the designs they build under licence.

According to the original version of the DSIR Report the failure rate of Diesel engines, to whatever design, built in British engine works was much higher than that of their competitors.

INNOVATION IN CONSTRUCTION METHODS

The main development in methods of construction has been the assembly of pre-fabricated components under cover. Although one of the very first covered berths was at Swan Hunters where the *Mauretania* was built in 1907 and at least one other British yard had covered berths, these were only roofed in against rain, and not enclosed; the sides being left open to the elements. Typically the British industry, having started an innovation, failed to develop it to its logical conclusion. The next stage has been the development of flow production methods leading, in the case of standardised ships, to a form of mass production. These methods were first used by the Americans during the war for the building of 'Liberty' ships, based on a British design for carrying dry cargo, but all-welded, and of tankers. Although there was much scepticism by traditional shipbuilders about these methods, in practice the ships, after some failures, were a success and have had reasonable lives. The Germans used similar methods to build submarines. The sections of the submarine were constructed in heavy concrete bunkers in Hamburg, fitted out with pipe work and electrical wiring, and then assembled by welding and the pipe and electrical connections made.

After the war these methods were taken up by the Japanese and then by Swedish and other yards; but British shipbuilders were slow to adopt them as they did not think they were feasible in peace time. Only one firm, Austin and Pickersgill of Sunderland, successfully designed and built a standard range of dry cargo ships as replacements for the Liberty ships. These were needed by the shipping group to which it belonged and their large requirement guaranteed the number of vessels with which to start the series at a competitive price. The design has continued to be successfully developed. Most British builders failed to appreciate the cost advantages of standardisation; such as the elimination of teething troubles, simplification of production methods and the spreading of the overhead costs of planning and controlling production. Recently there have been signs of change. In 1970 Court Bros. built the first fully under-cover flow production yard in this country; the Appledore Yard in Devon; but its capacity

is limited to ships of 10,000 tons; and Camell Laird are now developing a range of standard ships to be produced in new covered berths.

Another important innovation which started in Germany was the replacement of full-scale lofting, for the production of templates for cutting the steel plates, by one tenth scale negatives made by photographing the original drawings and projecting them enlarged to full size on to the surface of the plate to be cut. From this came a further development now in general employment: the use of negatives to control an automatic gas cutting machine. The next stage has been the use of computers in the design of the hull and the development of the outlines of the plates with finally the automatic cutting of the plates by a machine directly controlled by tapes produced by the computer. BOC/Ferranti, with DSIR support, designed and manufactured such a machine in the 1960's and two of them were installed in British yards; but the concept was too advanced and the programme too complicated to be commercially successful. This method of cutting is now common practice in Japan, Germany and other countries, and growing in UK shipyards. German cutting machines seem to have maintained their lead.

In general, British shipbuilders have come increasingly to fall behind their competitors, not only in their capital equipment and the lay-out of their yards but also in the application of management techniques such as production planning and engineering, production control, quality, control work study, operational research and cost accounting [2].

Vol. 4
1976

QUALITY OF MANAGEMENT

All reports have criticized the quality of management, especially in the middle ranks. The lack of attention to marketing has led to investment decisions seldom being based on a thorough assessment of the benefits in the market to be served [9]. It may also have been the reason for the failure to develop standard ships which need specialised selling [2]. The design effort has tended to be dictated by the short term demands of the customer and not directed to longer term developments. This was probably due to the lack of separate new product departments and the low ratio of qualified staff compared with those in the design departments of European yards: one in twenty-three compared with one in ten [2]. This has been especially important when dealing with advanced technology. The figures are difficult to compare; not least because while most professional engineers in other countries are educated in institutions of university level, this has not, until recently, been so in the UK. The DSIR report, in its published version, gave the number of 'graduates or equivalent' in the shipbuilding and marine engineering industries as over 1200; but the vast majority were members of the Institution of Marine Engineers; the qualification for membership of which was a Ministry of Transport (originally Board of

Trade) sea-going ticket. This was a qualification for operating machinery at sea; but hardly for designing and developing new machinery. In 1959, of new corporate members of some of the main engineering institutions the following numbers were graduates:

	New members	Graduates
Civil	1100	670
Mechanical	2680	670
Production	870	100
Marine	630	40

Since then the number of graduates has increased as the standards required for membership have been raised to meet the requirements of the Council of Engineering Institutions. In Japanese yards nearly every manager and engineer is a university graduate. At the time of the DSIR report it was estimated that the total number of graduates in the British industry was 80; half of them employed in YARD. Judging from the 1972 Report of the Shipbuilding Industry Training Board the numbers employed of qualified engineering staff of all kinds have been declining to about 60% of the 1965 figure; although the number of managers, administrative and clerical staff increased at a time when the number of manual employees had been falling by about a fifth. The industry has shown a lack of interest in training and few if any firms have serious graduate programmes [8]. This can be compared with the admittedly paternalistic Japanese companies which contrive to take the same number of graduates in boom or slump and to give them systematic practical training.

The industry's long record of bad industrial relations is well-known and it has led to low productivity and a decline in quality. Owners complain of long delays in delivery and lack of interest in quality control which is confirmed by the classification societies. The inflexibility of labour becomes an increasing handicap in a period of rapid technological change; but it may be that one reason why no attempt was made during the prosperous time after the last war to come to an agreement with the unions to 'buy out' the practices which often made the introduction of new processes unprofitable was the inability of management to foresee the technological changes which were coming. A vicious circle has thus been created of initial management conservatism and resistance of workers to change which in practice reinforces that conservatism. The end of the process is the unwillingness of better qualified and more entrepreneurial engineers to enter the industry or, if they do, soon to leave it for the classification societies or the ship brokers; so reinforcing its decline.

Vol.
197

RESEARCH AND DEVELOPMENT

Very little R&D is carried out by individual shipbuilders or engine builders (except GEC); but there has been a tendency for the larger owners to expand their own technical resources. The DSIR report stated that the total amount spent on R&D was insufficient and, in particular, no research was being conducted on production problems. The development of propulsion machinery was hindered by the organisational structure of the marine engineering industry. The criticisms of *Pamatrada* led to the hiving off of its research activities to the British Ship Research Association, which had been formed as the British Shipbuilding Research Association in 1944. In addition to the Association, research was being carried out by the NPL in its Ship Division and Marine Science Division, which is concerned with Communications and Navigation; in the Hydraulics Research Station; in the National Engineering Laboratory; in some Ministry of Defence Stations and by Lloyds. Co-operative research on Diesel engines was carried out by the British Internal Combustion Engine Research Association.

Vol. 4
1976

An increase in R&D took place in the 60's, although very little was carried out 'in house' by shipbuilders; but it continued to be directed into areas which would not give the industry particular competitive advantages [2]. The Rochdale Committee made some pertinent criticisms of the lack of co-operation between owners and builders on research programmes [4]. The Chamber of Shipping which, in 1967, had asked BSRA to put forward a programme of research to be sponsored by ship-owners came down against a joint programme with builders. The Committee, however, recommended an integration of existing R&D facilities to be supported by the shipping and shipbuilding industries; a recommendation which has never been carried out. On the other hand, some of the larger owners have carried out the recommendation that they should have strong technical departments responsible for all aspects of R&D carried out by, and for, the companies.

Areas in which the Committee thought further R&D was indicated were cargo handling, through transport studies, and ship manoeuvering and handling; studies on the first two of which needed close and continuing co-operation with those responsible for port studies and the disciplines they involve. They also wished to see more research on the basic problems of cargo transportation; which should be a major pre-occupation of the ship operator. This would include investigation of methods of optimising the complete through transport of particular cargoes from origin to destination, the interaction of ship and shore facilities and practices and so forth.

Since 1971 industry and the Government have reduced their contributions to BSRA. The Research Requirements Board of the Department of Industry, set up after the Rothschild Report, have replaced the block grant to BSRA by a

system under which support is related to specific programmes, with the emphasis on ship construction.

GUIDELINES FOR FUTURE POLICY

The Government has now assumed full responsibility for the industry's future policies. Whatever steps are taken to improve management and labour relations, there will have to be a policy on Research, Development and Design. The Geddes Committee reported against the industry's concentrating on 'sophisticated' ships; but, in view of the difficulty which most yards have in raising productivity to the levels of the industry's competitors, the best possibility of obtaining orders may well lie in seeking to supply particular markets and in offering an improvement in quality control and technical advantages in design. One of the disadvantages of not concentrating on a particular market, even if not building a standard range of ships, is that it leads to great variations in labour employed as each order requires a different balance of the various trades. If industrial relations are to be improved a building policy is needed that provides employment, as far as world conditions permit, for a stable labour force.

The emphasis now being placed on the systems approach should lead to an R&D programme co-ordinated between owners, builders and port managers; a recommendation of the Rochdale Committee which has never been adopted. It is on this approach that the arguments for large units with centralised management and technical services are based. Under present circumstances it would require extensive government support; for instance for the building of ships incorporating risky innovations as was done in Japan [10]. In the future the sophistication of ship design as well as the dimensions of ships will be increased by international agreements for the prevention of pollution at sea.

Subsidisation of this kind has not been very successful in the past in other industries, and it would be no cure by itself for the endemic weakness from which British shipbuilding has suffered for too long. It is certain that no action to revive the industry can have any hopes of success unless it provides a framework in which all its deficiencies: poor management, inflexible labour and technical backwardness, can be dealt with and this must involve an infusion of new managerial and technological blood and a radical change in Trade Union attitudes.

REFERENCES

1. Al Timimi W (1974) PhD Thesis, Exeter University, UK.
2. Booz, Allen and Hamilton (1973) *British Shipbuilding* 1972. Department of Trade and Industry, UK.

Vol.
197

3. BOWEN HG (1954) *Machinery and Mossbacks*. Princeton University Press, NJ, USA.
4. Committee of Inquiry into Shipping (1970) Report Cmnd. 4337, HMSO, UK.
5. Department of Scientific and Industrial Research, UK (1960) Research and Development Requirements of Shipbuilding and Marine Engineering.
6. DREWRY HP (1975) (Shipping Consultants Ltd) World Shipbuilding Output and Capacity.
7. HANSARD (1969) 3rd March, Col 42.
8. Productivity and Research in Shipbuilding (1972) Report of a Committee under the chairmanship of Sir James Patton, C.B.E., B.S.R.A.
9. Report of the Shipbuilding Inquiry Committee (1965–66) HMSO, UK.
10. YAMASHITA L (1967) Application of technical innovations and automation to ships built in Japan, in *Automation on Shipboard*, (Ed. BORWICK GJ) p. 57. Macmillan, New York.

ADDRESS FOR CORRESPONDENCE: *A. Albu, Esq.*, 17 *The Crescent, Keymer, Hassocks, Sussex BN6 8RB*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

OR, Organisational Design and Adaptivity[1]

ROLFE C TOMLINSON[2]

Received April 1976; in revised form June 1976)

The purpose of this paper is to discuss the need for fundamental research of an OR/systems nature into organisational design. This is illustrated through the work of a research team from the International Institute for Applied Systems Analysis which recently visited the Tennessee Valley Authority as the first part of a comparative study of the organisational problems of regional development. Emphasis is laid on the need for organisational adaptivity, defined as the ability of an organisation to adapt in the face of unpredictable external change. Attention is also given to the need for 'consistency', so that elements within the organisation are constrained to operate to the advantage of the whole.

Vol. 4
1976

'OPERATIONAL research is concerned with the improvement of decisions and decision making processes, both of which are inextricably entwined with the organisational system.' That is a statement of belief, but it is so fundamental to the rest of this paper that some preliminary discussion of it may be helpful. Clearly, so far as decision making processes are formalised, they *are* part of the organisational system. But what about the rather large number of major decisions taken in any organisation which are forced by external events, and for which no specific preparation has been made? Surely they can be considered in isolation, irrespective of the 'system'? Such a suggestion ignores the fact that in a good decision making body problems are *never* treated in isolation. The secondary consequences—political, economic and administrative—are all taken into account; the information used depends both on the formal information flow and the informal contacts the system provides. The decision taken is, more often than not, suggested from within the system, though there may have been prior suggestions as to the form the proposal should take. The ability of operational research to influence decisions is therefore strongly modified by the way in which the organisational system works. Additional confirmation of this fact comes from the well known fact that problems of deciding where to locate

[1] This paper was presented at the Second International Research Conference on Operational Research, at Stratford upon Avon, UK, 1976.

[2] Operational Research Executive, National Coal Board.

operational research within an organisation can almost invariably be traced back to fundamental inconsistencies in the pattern of responsibility within the organisation. So good organisational design is necessarily a matter of importance to those concerned with OR, as well as to all who work or live in organisations.

If, then, organisational design is so important to OR, and since it is clearly a systems problem in any case, one might have expected that OR people would expect to be involved in reorganisation problems, as part of the team if not directly in terms of an OR investigation. On the contrary, that seldom seems to happen. The reason for this is the subject of this paper.

At this stage, we need to satisfy ourselves that there is a need—that the state of the art of organisational design has not yet got to the point where expertise is soundly based and mistakes made infrequently. Consider, however, some of the most common reactions that one gets in discussing the results of a re-organisation.

1. Almost invariably, it is found that the organisation behaves quite differently from what is predicted. Its stated objective is often achieved, but that objective is often partial. Thus, a reorganisation aimed at reducing costs will usually manage to do so—it is a kind of self-fulfilling prophecy. But cost reduction cannot be the sole objective, else the sacking of all staff would be the complete answer. The organisation must have an output (profit, service, etc.), and this may be affected, often in an unmeasurable way. (As an aside how many organisations even *attempt* to measure the effect of changes?)
2. Tasks are designed independently of people. Similar units are given the same organisational system, regardless of the very different people who will have to fulfil them. This is particularly important with jobs of planning and co-ordination where the discretionary content is very high.
3. The dynamic effects of the change are often ignored. The shock wave of change strikes the organisation long before the changes formally take place, and the consequent ripples take a long time to die away. Change of this kind is very wasteful. An organisation that can adapt without catastrophic upheaval has a major survival advantage.
4. An increasing number of changes create a situation where the new system 'forces' subsidiary units to behave in a way that is inimical to the interests of the whole organisation. As an illustration of this, one only needs to point to the many organisations that have created service departments as independent cost centres, only to find that in so doing they have ended up with a greatly reduced level of service, albeit at lower cost.

It would not be difficult to argue the case that organisations are usually changed without means of telling what the result of the change will be; that too

Vol.
197

little attention is paid to dynamic behaviour of the system, and virtually none to the effect of change on those tasks which have a substantial creative or discretionary element.

If there is any truth in this, something needs to be done. But what? The research literature on the subject does not help very much, and when one comes to the crunch of a particular organisation the scientific investigator finds himself wanting to start from scratch, and he seldom has time for that. So, those concerned fall back on experience, usually of different situations and with inadequate feedback. In such a situation the OR/systems man—with his developed skills of analysing complex organisational problems—should have as substantial a part to play as most. But if he is to establish a firm claim to be consulted he must ensure that the necessary fundamental research is carried out.

In this paper I want to describe one attempt to do this which is being undertaken at the International Institute for Applied Systems Analysis (IIASA) near Vienna, and raise some of the issues and questions arising from that work.

THE IIASA EXPERIENCE

Vol. 4
1976

One attempt to establish such a scientific basis is at the International Institute for Applied Systems Analysis in Vienna. The Institute was established just over 3 years ago after a great deal of international negotiation, primarily between the USSR and the USA. It came from a desire by those two countries to find some area in which they could co-operate, in the mutual awareness that all developed societies—whatever their ideology—faced common problems, many of which cut across national boundaries and for which there were no established methods of solution. After much discussion these problems were seen to be systems problems, in the sense that they cut across most of the traditional areas of expertise and could only be studied by people experienced in the analysis of complexity and able to model the way in which such complex systems behaved. It was therefore agreed to establish IIASA, with the major funding coming from the USA and the USSR, though with a number of other countries taking part and supplying together a substantial small portion of the overall funding.

In the first place it was agreed that work should start in such fields as energy systems, ecological systems, urban systems, industrial systems, medical systems, etc. and a useful start has already been made. It was realised at the outset, however, that if the work of the Institute was ever to be applied in practice it would have to make an impact on the decision making processes within social organisations which might be political, bureaucratic or industrial. It was therefore essential that some research work should go on in the general area of planning and decision making in large organisations, partly to tackle specific problems, but also to ensure that the Institute had staff who could contribute personal expertise to the problems of the implementation arising elsewhere in

the Institute. It was therefore decided to establish a Large Organisations Project.

The next question was where, in this vast problem area, to make a start. One of the problems was of course the very different background and experience of those members of the team coming from different countries. Accordingly it was decided to make a comparative study of two similar organisations in the USA and the USSR. Because of Soviet concern with problems of regional development, which has since become a main feature of the work of IIASA, it was suggested that the first study should be made of the Tennessee Valley Authority (TVA) in the USA. This took place in July 1975 and is to be succeeded by a similar study of the Bratsk/Ilimsk region of Siberia in the summer of 1976. It is hoped that this comparison will yield direct scientific returns, but it would have been justified if it did nothing more than create a genuinely international team of systems scientists with a common experience in this area. I was fortunate enough, as an adviser to the project, to be a member of the team that visited Tennessee, and it was soon apparent that we were learning as much from the questions that other members of the team asked as we were from the answers, given openly and freely by the staff of the TVA.

The TVA is, of course, a most unusual organisation, both in the way it was set up, and in its powers. It was established primarily to control the waters of the Tennessee Valley by a programme of dam building, but the general purpose of developing the region has dominated its planning from the beginning. From the start, for instance, it has been advising farmers on how to improve productivity. It develops and manufactures fertilisers. It is concerned with forestry, recreation and industrial development. Above all, it produces power. Indeed the Authority is now one of the largest power producers in the world, with hydro-electricity only providing for about 20% of the total. 95% of the TVA's expenditure is on power, this being funded by normal commercial means. The remaining 5%, $70M, comes from Congress. The transformation of the Tennessee Valley over this period has been remarkable; yet the Authority possesses no direct powers, except with regard to flood control. Political control remains entirely with the seven states whose territory is included within the TVA region. The TVA is well documented in the literature and the report of the IIASA study should be available during 1976.

My particular assignment in the course of this study was to examine the present planning process at the TVA. This is relatively new—it has been in its present form for only 2 or 3 years—and has a number of features which are both of general interest, and also relevant to the theme of this paper. There are four main stages, as follows:

1. *The situation assessment*

In the autumn of each year, each division or office of the TVA prepares a 'situation assessment'. In this they discuss, from their own viewpoint, the

external situation facing the TVA, the possible impact of other divisions and offices within the TVA and, finally, the problems and possible solutions facing their own part of the organisation. In undertaking this assessment they are encouraged not only to identify trends and to state how they believe these trends should be met, but also to expose points of difficulty and potential conflict. All these assessments are considered by headquarters and discussed with the staff concerned. Once they have been approved they provide an agreed scenario in whose light the more detailed financial and technical planning can be carried out.

This is an effective procedure, avoiding much of the pre-emptive bidding and game playing arising in most organisations where forward planning gets confounded with competitive claims on available resources. It also has a remarkable therapeutic effect in clearing the air with regard to points of contention whose existence might otherwise be hidden from the Board until a more serious explosion occurs.

2. *Summary planning documents*
These documents are prepared at the beginning of the following year, projecting programme plans at least 5 years into the future. This is an office or division programme planning and policy statement, satisfying general goals, measurable objectives and dollar and man-year requirements for the entire planning period.

3. *The multi-year budgets*
These set out in greater, more specific detail, the manpower and financial plans for the same period. They form the basis of the TVA submission to Congress.

4. *Programme evaluations*
Each major programme is evaluated on a periodical basis, typically 3–5 years, so that the Board can review in detail the progress made in each subject area and can evaluate the work carried out.

One of the more surprising features of these procedures is the time the whole process takes. The final appropriation by Congress is made 2 years after the first situation assessment has been prepared! Such a delay would almost certainly bring the whole budget planning system into disrepute if it were not for two saving graces. One of these is that the money, once awarded, is 'no year' money i.e. if the programme cannot be completed within the expected 12 months the money is still available for use for completion of the projects. Secondly the money is awarded to the TVA as a lump sum and not tied to individual projects within it, so that some flexibility is possible. (It still has to be justified after the event.)

Internally the planning process seems to work well in most respects, though it is not equally suited to the work of all divisions and offices. It encourages

Vol. 4
1976

each division to plan ahead, explore alternatives, expose problems and review progress. In this sense the system is undoubtedly 'good'. The most obvious weakness is that very little effort is devoted to co-ordinating all this work at HQ. The staff is not available to conduct a detailed analysis, to compare and collate the various programmes or, for example, provide forward projections on a common basis. Certainly there can be no 'optimisation' of the whole process. Thus the planning process, although 'good', is certainly not 'efficient'. This proves to be an important distinction and we shall return to it later.

The TVA study was the first part of a comparative investigation, and was the first collaborative effort of the IIASA team. Ideally, more preparatory work should no doubt have been carried out to provide a more formal structure for the investigation, but without this common experience it is doubtful if the proposed structure would have proved to be of lasting value. This more systematic work is now taking place. But certain features stood out nevertheless which will, I am sure, colour the whole future of this work.

1. The first outstanding feature of the TVA was its flexibility. The Authority was established by Act of Congress in 1933 and the Act has not been altered to any substantial degree since then, though various features of it have been tested in the courts in such a way as effectively to extend their powers. Despite this, the activities of the TVA has altered very extensively over the 40 years of its existence. The most obvious change is the way in which power production dominates the finances of the Authority. Despite this, it has continued to operate effectively and overall to maintain the same ethos, that of a public service R&D organisation, thoughout.

2. The second main point was the need to distinguish between 'goodness' and 'efficiency' in organisations. The first requirement of an organisational structure is that it should achieve the desired objectives and motivate the staff to behave so as to further the objectives of the organisation as a whole. This can happen in an organisation which is not optimised in any sort of way. Equally it is possible to develop an organisation where each part is (sub-) optimised, but which does not cohere as a whole.

3. The study showed a most important feature that 'money need not talk'. A study of the financial structure of the TVA would have left one in little doubt that it would be dominated by its Office of Power; that the other functions would be subservient to the activity which generates 95% of the Authority's revenue. This does not happen for reasons which may be fortuitous and are certainly not fully explained. Part of it no doubt lies in the fact that, although its efficiency as a power producer has always to be justifiable under criticism, the Authority as a whole is not thought of as a 'business', nor is it expected to behave like one. Its behaviour pattern is that of a public service R&D corporation.

4. Above all, in my mind at least, it brought a realisation that a formal comparison of structures would ultimately be of little value. What was needed was an understanding of the organisational dynamics in response to changing circumstances—the problem of organisational adaptivity.

ADAPTIVITY

Organisational adaptivity is the ability of an organisation to change itself, or the way in which it behaves, in order to survive in the face of external changes which were not predicted in any precise way when the organisation was designed. I do not include within the definition the formal control processes built into any properly designed organisation, which enable it to cope with the anticipated perturbations in the environment.

Let me explain by means of an example taken from the IIASA organisation itself. Like all research organisations its ability to undertake work depends on the availability of the funds necessary to pay its staff. When IIASA was established methods were laid down by which the finances of IIASA were to be controlled and by which member organisations would pay the appropriate dues. Provision was made for a 20% increase in these dues after 3 years. What no-one had envisaged, however, was the fact that in the first 3 years of its existence there would be heavy devaluation of the American dollar, nor that there would be such an enormous increase in the rate of Austrian, and indeed world, inflation. Because of this the national dues were fixed in terms of dollars, and the 20% allowance for inflation has turned out to be derisory. Consequently IIASA did not have the money they expected to have and the short-fall is a great deal more than the 5% or 10% which one would expect to have to cope with in modifying a research programme. This has meant a fundamental re-thinking of the whole way in which the programme should be structured. Necessity is the mother of invention, but even so the re-thinking process has been slow and difficult. Originally the feeling was that any potentially useful idea was worth exploring. Now the work has to be concentrate on key issues, and many important topics deferred. Some staff have never really adapted. But the process is much more difficult when the external changes do not have such an immediate effect as lack of cash, but are nevertheless severe enough so that the future survival of the organisation requires changed plans. Adaptation is essential to survival.

Of course, the way that an organisation ultimately adapts is through the taking of decisions. Those decisions may relate to overall policy, to the way in which things shall be done in the organisation (i.e. procedures), or to more specific decisions as to whether a particular project shall be given the go-ahead or not. Although the critical decisions will be taken at the top (say, by the Board) one often finds that their freedom of decision is severely constrained by

Vol. 4
1976

the information and proposals put to them by their supporting staff. It is, in fact, necessary to make a clear distinction between 'decision making' and 'decision taking'. Almost invariably a board of management taking a decision does so on the basis of papers put to them by subordinate officials. These officials will have considered the pros and cons relating to a particular decision and will have submitted a relevant digest of these so that the decision takers (the Board) can formulate their judgement. Of course, the Board may ask for further information, or demand that proposals be modified. In the end they take their decision on the basis of other people's information, and no doubt choose the best from what they are offered. What, however, if they are not offered the right proposals? It could be that in the early stages of preparation the wrong alternatives were selected for examination—by ignorance, mistake or, as sometimes happens, selfish motives? If this happens the right decision cannot be *taken* because it has been *made wrong*.

A TVA example may help to clarify the issue. Shortly before the team's visit to Tennessee, the TVA had approved proposals for a nuclear power station at a particular site. The final decision they made was a simple one—do we go ahead or not? But should they have been considering a different site, or perhaps a fossil fuel station instead of nuclear? All these options were somehow eliminated before the final proposals were put forward. In fact, the most important decisions had in all probability already been taken, and some of them *not* by the Board.

This does not mean that the real power lies out of the Board's hands. Long before this proposal came along they had taken decisions on how such matters should be handled. But those decisions were probably more important than their choice of the final option. In the long run the Board needs to be judged on what it does to ensure that those decisions are *made* correctly. And if it is important, as surely it is, for an organisation to be adaptive, then the decisions about decision making procedures must take this into account.

One of the main elements in the decision making process, and certainly of prime importance when we consider questions of adaptivity, is planning. Organisational planning is a much more difficult concept than is always realised; indeed, protagonists of different planning systems often spend their time arguing about quite different things. Although I have already spent some time talking about the TVA planning system, it is important to realise that the supposed separation between planning and operation is more often confusing than it is helpful. The danger of this mythical separation is, of course, revealed by that simple question "After the plan fails, who is to blame—the planner or the operator?" The organisational separation of planning from functional control is probably necessary in most organisations, but a superficial examination of the internal dynamics of the situation will quickly show up the problems.

There is, however, one crucial factor that is too often omitted from discussions about planning. Planning is often discussed as if the preparation of a

plan is a process equivalent to the manufacture of a motor car. That is to say that it assumes that one can set up a procedure for feeding in standard inputs (usually in the form of statistics and other information) and that by appropriate manipulation and analysis one can produce a set of finished plans which satisfies the given criteria in some optimal way. On the contrary, although there are routine and mechanical elements in most planning departments, the essence of good planning is that it is an imaginative exercise, depending critically on the quality of ideas, the perception of need and the understanding of implications. These factors are just as important as the demands of formal logic and clear exposition. If, therefore, planning is to be adaptive, changes in the external situation must somehow be brought to the consciousness of the planners who are involved in this imaginative process (not just their notice). Only so will their thinking be modified accordingly. Thus we have a serious problem of intelligence and information.

At this point I must make an aside. Informatics experts will inevitably have been waiting for me to refer to communications and information systems, for surely these must have a major part to play in creating organisational adaptivity. That is certainly true and I have only touched on part of their role. But I must at the same time express two reservations. In the first place the kind of information that is needed to create awareness is often quite different from that which is provided by traditional information systems. It is not so much a question of statistics as of understanding. My second reservation is more serious still. Information systems, like control systems, are designed to operate and work within a predicted range of conditions. If needs occur lying outside the predicted range then they may become irrelevant or misleading. The danger with information systems is quite simply that they present information in a standard form which may lead one to ignore or to be unaware of important information which is simply not transmitted because it lies outside the standard pattern. My concern here is not academic—it is a commonly voiced concern of managers in many organisations. Seen from the inside of an organisation, computerised information systems often seem to have a massive rigidity combined with substantial inbuilt obsolescence.

Before going on to the final section of this paper, which is concerned with the need for further research, it may be helpful to summarise the main threads of my argument so far.

'The effective survival of an organisation, the ability to meet its prime objectives in a changing environment, is defined as organisational adaptivity—it includes what we would more commonly describe as flexibility. The ability to adapt depends to a large extent on the way in which decisions are made (which is a distinct concept from the way in which decisions are taken). A key element in adaptive decision making is the role of the planner, and his most important contribution is creative—it cannot be prescribed. How he performs, and whether he enables the organisation to adapt, will depend on:

Vol. 4
1976

(i) his awareness of the changes in external circumstances, and of shifts in policy associated with them; and

(ii) the internal pressures and controls that are exerted on him by the organisational system and which can easily force him to behave against the overall interest of the organisation.

TACKLING THE PROBLEM

Having defined the problem we now have to decide how to tackle it. I have already said that I do not think that existing work takes us very far, though it is useful in providing concepts. So research is needed. What kind of research? Well, progress in a scientific manner can only be made by the repeated application of the following steps:

1. Classified observation.
2. The development of theories or hypotheses.
3. Measurement or experiment in order to test those theories or hypotheses.

Let us look at how some of the IIASA team are trying to tackle the problem.

At the start of the IIASA work it was agreed that because of the lack of firm concepts on which to build, and in order to create a common body of experience between the members of the team, two organisations should be examined in depth. Three others are also being examined in less detail. A group of young scientists from Russia, Poland, USA, UK and Bolivia have been trying to develop a common basis for description, measurement and analysis of the (very different) organisational structures. Their basic argument, somewhat rationalised by myself, is as follows:

(a) They are not seeking a single organisational ideal (organisation is in fact not a single concept), nor is the purpose to make value judgements of one against another.

(b) The analysis must be structured on a common framework covering the major activities of the organisations. This framework must be culture-free.

(c) The best way of achieving (b) is by examining operational functions. However, the traditional departmental-type functions used in organisation charts are not appropriate, and it was decided to use the following five main categories:

1. Policy making.
2. Development (External planning & Research).

Vol.
197

3. Group management.
4. Co-ordination (Internal planning and control).
5. Operational management.

These functions have the advantages of recursion—i.e. they appear at all levels of an organisation.

(d) The question of measurement remains unsolved. Basically it needs, in a very general sense, to be viewed as the degree to which achievement corresponds to intention—for every function, at every level. But an important intermediate concept of 'consistency' is also introduced, which basically is concerned with ensuring that the internal controls ensure that intermediate elements do the right kinds of things. The concept is virtually the same used when discussing the 'goodness' of planning systems as compared with their 'efficiency'.

(e) Measures of static performance are not adequate, since dynamic performance is likely to be critical. This requires the follow through of critical decisions from the germ of their conception and the study of certain control situations. Some kind of sensitivity is also implied.

Vol. 4
1976

If the results are to be co-ordinated, and indeed if compatible data are to be collected, some basic theories or hypotheses must be used. Consideration is being given to Beer's Model of the Firm, a theory based on field force theory in physics, and Lewin's model of planned change. None of these look at present to be quite what is required.

CONCLUSION

The purpose of this paper has not been to propose a new theory of organisational design, but to suggest that this is an area of importance that still needs a lot of research—and that this needs to be done in the applied, interdisciplinary mode of particular concern to the OR/systems man. It is hoped it will stimulate response and a fruitful exchange of ideas.

ADDRESS FOR CORRESPONDENCE: *RC Tomlinson, Operational Research Executive, National Coal Board, Coal House, Lyon Road, Harrow, Middx HA1 2EX, UK.*

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Government Policies Towards Innovation: A Review of Empirical Findings[1]

K PAVITT

(*Received March 1976*)

Since the Second World War, governments in Western mixed economies have become increasingly involved in promoting technical innovation in industry. The analytical basis for this involvement is deficient, as is empirical information and analysis on its effects. But it is probable that, from an economic viewpoint, governments have been mistaken in supplementing normal industrial activity by financing large-scale development projects in technically sophisticated industry. They should instead be financing longer-term basic and applied research, as well as increasing technical awareness in more traditional industries. More needs to be known about the factors affecting innovative decisions in industrial firms, including the effects of industry-specific characteristics and of government policies.

Vol. 4
1976

SINCE the Second World war, governments in Western mixed economies have been increasingly involved in, and concerned about, the processes of technological innovation in industry.[2] The fact of government involvement is shown most clearly through the civilian research and development (R&D) activities that are financed by governments for the purpose of promoting innovation and its diffusion in the economy (see Tables 1 and 2). The reasons for governmental concern are more complex, but amongst them have been the results of economic research that has tended to confirm what the classical writers said in the 19th century, namely, that technical innovation and its diffusion are one of the main springs of economic growth and industrial competitiveness [35, 47].

However, economics has made a less notable contribution to public debate about the appropriate form that the government involvement should take. A number of reasons have been advanced to explain this state of affairs: the

[1] This paper is based on one prepared for the Royal Economic Society's Conference on Government and Innovation, which was held at Pembroke College, Cambridge, from 14 to 17 July 1975.

[2] For the purposes of this paper, we shall define technological innovation as the technical, industrial and commercial steps which lead to the marketing of new and improved manufactured products and to the commercial use of new and improved production processes and equipment. It should be noted that this definition of technological innovation is different from 'technical progress' as normally defined by economists. The latter refers to increases in the productivity of labour, capital and land. It therefore comprises not only technological innovation but also the diffusing of these innovations amongst their potential population of users, as well as managerial and organisational sources of increased efficiency.

TABLE 1. GOVERNMENT EXPENDITURES ON CIVILIAN, INDUSTRY-RELATED R&D COMPARED TO INDUSTRY'S OWN R&D EXPENDITURES

Country	Year	Industry-financed R&D as % of GNP	Government civilian, industry-related R&D as % of industry's own R&D expenditures	Same as previous column for government R&D outside the high technologies
Belgium	1971	0·6	31	8
Canada	1971	0·4	97	n.a.
France	1971	0·6	58	7
FR Germany	1971	1·1	26	3
Italy	1971	0·5	28	7
Japan	1971	1·2	8	n.a.
Netherlands	1971	1·1	15	4
Norway	1970	0·5	27	n.a.
Sweden	1971	0·9	12	n.a.
UK	1970	1·0	31	4
USA	1971	1·0	41	4

Government R&D Expenditures in the second column include those in the OECD's 'Mining and Manufacturing', 'Nuclear', and 'Space' categories of Government R&D expenditures. Those related to defence, agriculture, services, health, welfare, pollution, and basic research are excluded.

Source: Fabian, Young *et al.* Reference [13].

economist's ignorance of the workings of science and technology; the difficulties of making meaningful, quantitative measurements; and the persistence and elegance of theories of the firm that treat innovation as an exogenous variable [16, 24, 34]

A growing number of empirical studies of innovation in industrial firms—undertaken by economists and by scholars from other disciplines—have increased our knowledge of the nature of industrial innovation, and of the conditions associated with its success and failure. But for the formulation of government policy, some important gaps remain. We have little first-hand evidence about the reasons why firms do, or do not innovate, and about the influence on firms' innovative decisions of factors on which governments have some influence. Furthermore, governments have been more reluctant than industrial firms to open their books, hearts and minds for the critical review of their policies for innovation by independent scholars, or to undertake public evaluations themselves.

Nevertheless, we do not have to start completely from scratch. Given the considerable expenditures of public money made for the purpose of promoting industrial innovation, some economists have cast aside the neo-classical theory of the firm, and have tried to develop ground-rules to guide, or against which to evaluate, the nature, extent and purpose of government involvement in industrial innovation [6, 12, 25, 32, 33]. In addition, our knowledge of what

TABLE 2. GOVERNMENT SPENDING ON CIVILIAN, INDUSTRY-RELATED R&D, 1971

Country / Sector	Belgium 10^3 BF	%	France 10^6 FF	%	FR Germany 10^6 DM	%	Italy (1970) 10^9 L	%	Netherlands 10^6 FL	%	UK 10^6 £	%	U.S.A. (1973) 10^6 $	%
Aircraft	10	0	500	16	198	8	0	0	23	9	95	51	306	8
Space	501	16	638	20	516	20	8	7	45	18	12	7	2680	67
Nuclear	1747	56	1363	42	1224	49	59	51	108	43	50	27	607	15
Computers and automation	52	2	314	10	249	10	20	18	7	3	2*	1	0	0
Other industry	820	26	390	12	330	13	28	25	69	27	26	14	409	10
Total	3130	100	3205	100	2517	100	116	100	252	100	185	100	4002	100

* In 1972/3, UK Government R&D expenditures in the category computers and automation had increased to £12 million, largely because of a new loan to International Computers Ltd.

Source: European Communities, *Public Expenditure on Research and Development in Community Countries, Analysis by Objectives*, 1969–1973, 1974. National Science Foundation, *An Analysis of Federal R&D by Function*, Washington, 1974.

government policies actually are has also grown as a result of official reports from national governments, from the publications of the OECD, UNESCO, and the EEC, and from the collection of statistics on government and industrial expenditures on research and development activities. Critical evaluations of various government policies towards innovation have been made by independent scholars [3, 19, 22, 51]. Finally, some government bodies have actively encouraged and participated in the carrying out of such evaluations.[3]

Thus, whilst it would be misleading to talk of a coherent literature on government policies towards industrial innovation, relevant information, analysis and hypotheses can be dug out from a variety of disparate sources. This makes a survey of this sort more difficult, but perhaps also more necessary. No attempt will be made here to cover all sources exhaustively.[4] Instead, I shall try to select the most important findings, and to identify important gaps in empirical knowledge, within the framework of possible economic justifications for government interventions in industrial innovation.

2. RESEARCH AS A PUBLIC GOOD

Economists have argued that there will be a tendency towards underinvestment by industrial firms in basic research [32]. Since it is uncertain if and where a piece of basic research will be applied, and since it will often be applied in combination with the results of other research, the system of publishing the results of basic research increases the likelihood that practical applications will be identified. As a consequence, however, industrial firms will not be able to pre-empt all the benefits of the basic research that they undertake, so that without government funding there will be a tendency towards underinvestment.

Either explicitly or implicitly, government behaviour in all industrially advanced countries reflects such an economic argument, since governments are the major source of funds for the performance of basic research. However, the economist's tools of cost/benefit analyses are very blunt instruments with which to attempt any calculation of the 'optimum' level of government or national expenditure on basic research. Given that basic research gets applied in combination with other knowledge and with development, production and marketing activities, the attribution of any meaningful figure of economic and social benefit to past basic research is very difficult [5]. Given the uncertainties

[3] In particular, it is worth mentioning the Swedish Government's encouragement of a study of the development of the military aircraft *Viggen*, and the US Government's encouragement of a study of the development of the *Polaris* submarine [11, 43]. In France, the Délégation Générale à la Recherche Scientifique et Technique (DGRST), and the Ministry of Industry are financing a series of studies of French Government policy towards industrial innovation. In the U.K., a paper which evaluates (very critically) the Government's experience in giving launching aid in the aircraft industry, and which was written by an economist in the Department of Industry, has recently been published [17].

[4] This has been attempted elsewhere [40].

about cost and outcome, the *ex ante* assessment of the costs and benefits of proposed avenues of research is well-nigh impossible.

Some governments have therefore tried to use levels of expenditures on basic research in other countries as a guide to national resource allocations. For example, it has sometimes been argued that, UK expenditures on basic research (and especially academic research) has been "too much" by international standards. Unfortunately, the facts do not bear this out. In 1970/71, expenditures on academic research in the industrialised OECD countries were typically between 0·2 and 0·4% of GNP, and those in the UK amount to 0·2% [13].

There remain other important analytical and policy questions in addition to those of overall resource allocations: in particular, to what extent, where and how does basic knowledge get applied in industry? A number of empirical studies have addressed themselves to these questions, and their results can be summarised as follows [2, 18, 23, 30, 42]. In-house innovative activities are not the only sources of knowledge inputs into industrial innovations; between one- and two-thirds of these inputs came from outside innovating firms, and about half these extra-mural inputs came from universities, and from government-funded laboratories. These institutions play what is an essentially supportive role to industrial innovators, providing them with theoretical background, and with specific knowledge and information. The extra-mural knowledge is transmitted into the innovating firm through personal contacts and through the written word, and is often not reflected in financial transactions. Knowledge transmission tends to be greater in big than small firms, and in reasearch-intensive industries than in traditional ones.

Vol. 4
1976

These results have implications both for further analysis and for policy. Analytically, they suggest that economist's 'public goods' arguments for government support of basic research should be extended to include some of the basic and applied research activities of government laboratories. They also suggest the need for a more systematic and detailed taxonomy of industrial products and processes, in the manner in which they interact with basic and applied research activities.

For policy, they suggest that the general drive to reduce the size of government laboratories, which has been prevalent in many OECD countries over the past 10 years, may be dangerous if it ignores the type of contribution described above, and if one accepts the arguments (described later in this paper) that it is the proper role of government to *complement* the innovative activities of industrial firms by financing long-term basic and applied research, rather than to *supplement* them by financing short-term, commercial development activities. But there remains the problem identified by sociologists of maintaining efficient two-way communications between government laboratories and industrial firms [3].

Finally, there is one element in the government-financed scientific and technological infrastructure in Britain that has received particular attention,

namely, the low status of engineering. This is reflected in the quality of students opting for training in engineering, in their subsequent career pattern, and in low place held until recently by engineering research in the financial and academic pecking order of the Science Research Council. The role of engineers and engineering in Britain is in stark contrast to the situation in countries where engineering education is based on the German model, including the Netherlands, Sweden and Switzerland [15]. Britain's poor performance in world markets in engineering products has been attributed to this difference, but there is not yet any systematic empirical evidence to substantiate this view.

3. RISK AVERSION AND SHORT TIME HORIZONS

In addition to the 'public goods' reasons for government support of basic and applied research, economists have suggested other reasons why industrial firms tend to underinvest in long-term, more risky and radical innovative activities, namely risk aversion and short time horizons.

Freeman [16] has documented the empirical evidence that industrial R&D activities tend to be focused on short-term improvements. He argues that greater use of discounted cash flow methods in evaluating R&D projects will increase the pull towards the short-term, and he indentifies five classes of firm likely to accept a high degree of uncertainty in innovative investments: (i) government-subsidised firms undertaking high-risk ventures because of government acceptance of grossly optimistic assessments of future returns; (ii) government-subsidised firms undertaking high-risk ventures because of urgent national needs; (iii) small firms and individuals with strong personal commitments to particular inventions and innovations; (iv) firms constrained through their unfavourable competitive position to take big gambles; (v) large firms who deliberately put aside part of their funds for risky ventures, not closely controlled by formal methods of project selection.

It can be assumed for the purposes of this paper that the first category of high risk innovations, although regretably real, is not to be desired. It can also be argued that the fourth category is sometimes inevitable, but that it should not necessarily be given special encouragement when the welfare of very large numbers of people depend for their livelihoods on the outcomes. As to the fifth category, it includes in one sense the significant proportion of large firms that undertake longer-term and more speculative basic and applied research, often in central research laboratories established specifically for this purpose. Such firms are willing to undertake longer-term research provided that they can spread their risks over a sufficiently large number of relatively low-cost projects. However, we know that no firm was willing or able to undertake the costly, risky and long-term research in the early stages of the development of nuclear

Vol. 197

fission reactors, just as no firm today is willing or able to do the same in fusion research. This suggests that large firms' reluctance to do certain types of long-term research is not just a matter of aversion to risk or to the long term, but also a matter of cost. If this is the case, it would be useful to obtain more information on implicit or explicit trade-offs amongst cost, time horizons and risks in industrial firms' long-term research programmes, in order to build a somewhat firmer basis for government decisions as to the type of longer-term applied research that it should support.

4. TECHNICALLY BACKWARD MANAGEMENT AND INSTITUTIONS

Some economists have argued that industrial firms may also tend to under-invest in all forms of innovative activities to the extent that the institutional base of the industry and its management procedures became established before the potentialities of R&D became apparent [33]. It is for this reason that governments in all OECD countries have established comprehensive research, demonstration and advisory services in agriculture—an industry that can be characterised by the existence of a large number of small firms, and by innovations mainly in production methods rather than in products. And for similar reasons, governments in most Western European countries (but not in the USA) have established research associations (RA's) to perform R&D and related activities relevant to a whole industry (e.g. textiles), or to techniques relevant to a number of industries (e.g. welding) [26, 36].

Vol. 4
1976

Most of the activities of the RA's are outside the research intensive industries (i.e. outside the aerospace, electrical and chemical industries), and their relative importance in an industry's total R&D increases as the R&D intensity of the sector decreases. The pattern and purpose of their activities is similar to government programmes in agriculture: basic and applied research, coupled with information and advisory services for the purpose of improving and diffusing production and process technology, and of upgrading the technical awareness of management. In nearly all countries, the RA's comprise the one government programme of industry-related R&D expenditures where the allocation of resources is skewed towards the traditional industries with low research intensity. As Table 2 shows, this is in sharp contrast to the totality of such governmental R&D expenditures, which are skewed towards a few, privileged 'high technologies' (aircraft, space, nuclear energy, computers and automation).

To this author's knowledge, no cost–benefit analyses have been published on the work of any of the RA's, although a number of less rigorous measures suggest their usefulness. In a recent French survey, industrial firms identified them as the most frequently consulted source of extra-mural scientific and technical information, and similar conclusions have emerged from some British studies

[40, 42]. Similarly, the growth in the British RA's contract income from industrial firms can be interpreted as reflecting their success in improving management's technical awareness. At the same time, most RA's have not succeeded in reaching and influencing a large proportion of small firms with less than 200 employees, and the RA's advisory services are not on a scale—equivalent to those in agriculture—that would enable them to do so [7].

The proportion of government's industry-related, civilian R&D expenditures devoted to the RA's remains very small. In all Western European countries, it accounts for a relatively small proportion of the already relatively small 'other' category in Table 2 (about 20% of it in the case of the UK). Furthermore, it is significant that the British RA's get only one-third of their income from Government, whereas the Government provides 100% of the income of its aerospace laboratories. If lack of technical sophistication of the industry were the main guiding principle for the degree of government R&D subsidy, one would expect exactly the opposite pattern.

5. INAPPROPRIATE ECONOMIC REWARDS AND INCENTIVES

Some economists argue that tendencies towards short-term and safe R&D in some industries, as well as general technical incompetence in others, are a consequence of an inappropriate set of economic rewards and incentives for innovation, and can be overcome by their manipulation. However, to this writer at least, it would appear that government's manipulative power is very small, given the lack of firm evidence about the way in which various rewards and incentives—competition, monopoly, patent protection, profits and taxes—influence the firm's propensity to undertake R&D and to innovate.

For example, empirical studies by Grabowski and Scherer have come to conflicting conclusions about the influence of financial slack on the firm's propensity to innovate, and more generally, economists continue to disagree about the effects of monopoly and competition on innovation [44]. Freeman, on the other hand, concludes that, given the big uncertainties in innovative investment, what Keynes described as "animal spirits" has a considerable influence on the firm's propensity to innovate; for the same reason, he concludes that the process of competition through innovation will sometimes weed out the unlucky, as well as the inefficient [16].

What appears to be lacking is systematic, first-hand knowledge of the factors affecting the firm's decisions to innovate, or not to innovate, including factors related to economic rewards and incentives. Nonetheless, some such knowledge does exist. In a recent enquiry by AD Little [27], 24 US industrial firms were asked to identify what they perceived as the most important barriers to innovation. By far the most important were those over which governments have little

or no control: uncertainty over customer and market response, inadequate information for thorough evaluation of innovative ventures, and a management reward system that does not encourage risk-taking. The only area of government action that was perceived to be a significant barrier to innovation was uncertainty about Federal regulatory policies (the enquiry was undertaken at a time of great public and political concern about environmental protection). However, this particular enquiry has a number of limitations, related to the small number and heterogeneity of the firms sampled, and to the restricted range of questions asked. In particular, no questions were asked about the way in which industrial firms would respond to changes in government policy.

This is one of the significant contributions made by Taylor and Silberston [46] in their work on the patent system. In particular, they asked UK firms how they would react in their R&D expenditures to a government policy encouraging much more actively the licensing to other firms of patented inventions. The response was that in the drugs and fine chemicals sector, such a change would lead to a considerable reduction in R&D expenditures, whereas in most others it would make very little difference.

This result suggests that different things might be important for innovation in different sectors, depending in large part on their technical and institutional characteristics. Nelson and Winter [34] have pointed to the big differences in technical innovation in commercial aircraft and in agriculture. In the former, firms can expand capacity rapidly, and it is costly and time-consuming to imitate somebody else's successful product. In the latter, it is difficult for a successful firm to expand rapidly, and imitation is relatively easy. One can add that, in the case of drugs and fine chemicals, innovation is costly, imitation is easy, and that successful innovators and (in the absence of patent protection) imitators can expand rapidly. These observations require more rigorous formal development and empirical validation. Their public policy implications are obvious if not very helpful, namely, that different things may need to be done in different types of industry.

However, some economists have argued that one factor of importance to invention, innovation and diffusion in all industries is the overall level of economic activity [33]. Schmookler has shown how, in the USA in the 19th century, the level of inventive activity (as measured by the number of patents) in various industrial branches closely followed levels of investment activity. For a much more recent period, Falk [14] has shown a close relationship in all sectors of US industry during the late 1950's and the 1960's between trends in industry-financed R&D and trends in industry sales. This confirms at the macro-industrial level the findings of Seiler [45] in over a hundred US firms that top officials tend to use the R&D/sales ratio as one of the main guides in fixing and assessing the level of R&D activity. The same phenomenon may explain the relatively more rapid growth of industry-financed R&D over the past 15 years in continental European countries, by comparison with the UK.

Vol. 4
1976

Other recent studies have shown the importance of production and market requirements as stimuli to innovations [16]. And the one international comparison of rates of diffusion of industrial innovations concludes that one of the most important factors explaining differing rates is variation in trends in investment-inducing demand [31]. All of which suggests that continuing growth ensures the continuing increase in the resources and the opportunities for invention, innovation and diffusion, thereby laying the basis for further growth.

6. EXTERNAL COSTS

It is well known that the market mechanism may not reflect the full social costs of existing and new technologies. It is for this reason that governments have undertaken R&D in the past in relation to, for example, food, drug and transport safety, and to industrial health and safety. It is for the same reason that there has been considerable public pressure in the past few years for more government R&D activities related to environmental preservation.

Although such R&D has increased, it still remains very small indeed when compared to other items of government R&D expenditures [13]. There are several possible explanations for this: (a) governments are not doing enough; (b) governments are doing the less costly background research, whilst industry is making environmentally desirable technical changes induced by government taxes and regulations based on its background research; (c) environmental research is cheap compared to other forms of R&D financed by governments. No detailed empirical research has been undertaken to test any of these explanations.

7. IMPERFECT USERS' KNOWLEDGE

All empirical studies show the importance of user requirements in stimulating new directions for industrial innovation, and of attention to user requirements as the most important factor distinguishing successful from unsuccessful innovators [15]. In other words, the studies do not tend to confirm the Galbraithian attack on consumer sovereignty, quite the contrary.

But these empirical studies have their limitations. They concentrate almost exclusively on innovations in producers goods, and include hardly any in consumers or public service goods. This concern with producers goods in the first round of empirical studies can be justified because most R&D activities are concentrated in such goods. Nevertheless, as Freeman has pointed out [16] users of producers goods generally have the technical competence to specify what they need or to evaluate what they get, and they have to live with the consequences of the purchasing decisions that they make. These conditions do

not necessarily hold for consumers goods, or for certain classes of public service goods. The influence of advertising, packaging and incomplete or misleading information on consumers goods purchases is well known. There are also documented cases—the purchase of computers for the 'management' of local government in the USA, and the building of high-rise flats in the UK—where the purchase of public service goods have resulted from fashion and pressure rather than from careful appraisal of user needs [28, 49].

More systematic study is required of the factors differentiating success from failure in real technical innovations in consumers and public service goods. But even in the absence of such study, an economic argument can be made that governments should help provide 'countervailing technical power' to the purchasers of such classes of goods. Yet in no Western country does 'consumerism' appear to have emerged as an explicit function for government financed R&D.

8. SCALE OF DEVELOPMENT

Vol. 4
1976

Economists like Galbraith have argued that the absolute scale of effort required for the development and commercialisation of certain technologies can be mounted only with government money. Such an argument has been used in many OECD countries to justify considerable government support for the aircraft, space, computers and nuclear energy sectors (the so-called 'high technologies'). As Table 2 shows, it is these high technologies that pre-empt the lion's share of government funds for civilian, industry-related R&D activities. At the same time, the supposed inadequacies of the commercial capital market in supplying funds for the commercial development of innovations, that is implicit in this justification for government support, has also been used in relation to the establishment of government-financed institutions like the National Research Development Corporation (NRDC) in a number of Western European countries, in order to finance much less expensive commercial development outside the high technologies.

Over the past 20 years, economists have shown considerable interest in this line of argument; first by looking in detail at the relationship between size of firm and innovative activity, and second by questioning explicitly the economic desirability of certain forms of government support. In examining the relationship between level of R&D expenditures and size of firm, they have found that, over an employment range of up to five to ten thousand, the proportion of firms performing R&D grows with increasing size [16, 44]. However, studies in the USA have shown that amongst the largest firms, the proportion of resources devoted to R&D often does not grow with increasing size, and in some cases even decreases. Few studies of a similar nature have been done in Western

Europe.[5] Yet one justification for the encouragement given by many Western European governments in the late 1960's to mergers amongst large companies was the existence of a positive relationship between firm size and the relative level of R&D activities [48]. Once again, however, one must bear in mind differences amongst industrial sectors. A number of studies have shown a much greater relative contribution to R&D and to innovation by small firms in the electronics, scientific instruments and machinery industries than in large-scale industries like chemicals and aircraft [16].

More recently, some economists (and a political scientist) have argued government financial support for commercial development is not economically justified [6, 12, 22, 25, 51]. Industrial firms are normally capable of spending large sums of their own money on innovations as they see the strong likelihood of a reasonable rate of return (see, for example, the chemical industry and the technology for getting oil out of the North Sea). As a consequence, they will ask for government money for what they consider to be 'second best' projects. And there is no reason to believe that, outside traditional sectors, government officials and politicians can guess commercial prospects better than industrial firms, since governments will depend on industrial firms for information basic to evaluation, and (above all) since they cannot know very much about the all-important user requirements.

The argument continues that government financial support for commercial development will have a number of degrading effects, in addition to wasting money. Once started, government supported projects will be more difficult to stop than those supported with industry funds, because it is easier to spend other people's money, and because of political pressures. Government officials will be constrained to support commercially questionable projects in public. Industrial firms will devote their energies to political lobbying instead of to production and marketing. And scarce managerial and technical resources will be diverted from commercially more promising projects. At least some qualitative indications can be seen of these degrading effects: for example, the diversion of high quality scientists and engineers into aircraft and other high technologies in the UK since the Second World War [41].

But the economists' argument stands or falls on the assumption that industrial firms ask for government money for second best projects. No studies have yet been published which examine first-hand the characteristics of the commercial development projects for which industrial firms seek government money. Given the scale of resources involved, this is undesirable, if not surprising. This author has calculated elsewhere [29] that, of the British Government's expenditure on civilian, industry-related expenditure on R&D (excluding nuclear), about three-quarters is spent on commercial development, and about half on the commercial development of airframes and aircraft engines. The financial record of such government support is an indication *ex post* of whether projects

[5] Although one that has comes to similar conclusions [29].

have in fact been second best. In this context, one recent study by Gardner [17] is particularly revealing. Since the Second World War, the British Government has recovered less than one tenth of its expenditures on launching aid for aircraft and aircraft engines. It is to be hoped that similar studies will be undertaken for other programmes and other countries in future. One unpublished study in France also throws doubt on the scale justification for government support for commercial development: most of the funds spent on development aid ('aide au developpement') have been for *small* projects in *large* firms [37].

Whilst the available evidence suggests that government support for commercial developments is not economically justified, a case has been made by Eads and Nelson [12] for selective government support for long-term pure and applied research, and exploratory development activities, in areas where scientific and technical development may be too slow from an economic or social point of view. They argue that there are several possible reasons why this might happen. For historical, institutional or financial reasons, private initiatives may not enable a technology to be put on a sounder scientific basis, or sufficient attention to be given to exploratory development activities to test the broad attributes of new product or process designs; alternatively, because of market imperfections, governments may decide that a new technology needs to be developed faster than the pace established by industry. In such circumstances, governments should support applied research and exploratory development in order to speed up scientific and technical development. By comparison with commercial development activities such programmes would be relatively cheap. They would cover a broad front and should be managed on an incremental, step-by-step basis with the purpose of reducing key scientific and technical uncertainties to a degree that private firms can use the resulting knowledge to decide when (with their own money) they should move into full-scale commercial development.

Eads and Nelson argue that this approach has been successfully followed in the USA by the Department of Agriculture, by the National Advisory Committee on Aeronautics (NACA) in the 1920's and 1930's, and by the Atomic Energy Commission (AEC) in the 1950's. It has also been followed since the Second World War by the Japanese Government in shipbuilding.

If they are right, one can only conclude that British Government machinery is sadly deficient in one important respect. Outside the high technologies, there is no institution or formal source of funds for such exploratory work. For example, Sydney Jones eventually obtained government money for the development of British Rail's Advanced Passenger Train through the scientific old-boy network in the Government. One cannot be sure that the old boys in Government will always be so far-sighted and so influential.

Finally, mention must be made of the problem of the appropriate criteria for selecting areas for such long-term support. Whilst it can be argued that the economist's tools are appropriate and should be used very stringently in

Vol. 4
1976

evaluating any proposal for government support for commercial development, the same is not necessarily true, for longer term, exploratory projects. Indeed, it can be argued that the economist's criteria may lead to the selection of short-term projects [40]. The place of economic considerations in the selection of long-term projects is a subject that deserves further clarification. One place to start would be the industrial firm, which makes such decisions all the time.

9. SECURITY OF SUPPLY

Economists have been influential in criticising the strongly autarkic tendencies in government support for industrial technology, by pointing out that no one country can hope to create all the technology it needs, and that for most countries the introduction and diffusion of foreign technology is more important for economic growth than that developed nationally [50]. Nonetheless, international security of supply remains a powerful component of government policy for technology. Heavy government support for commercial development activities in the high technologies has been justified also on the grounds that it concerns strategic technologies that are crucial to the functioning of the modern industrial system, and that it is therefore economically dangerous to become critically dependent on foreign sources of supply, and on foreign technology, when these sources are capable of exerting monopoly power. Such arguments were prevalent in Western Europe in the 1960's when there was concern about the 'technological gap' between the USA and Western Europe [7]. They have re-emerged in the past 2 years in the wake of the oil embargo and the higher prices of energy and raw materials.

The fact of monopoly power cannot be denied. The US Government holds it in relation to nuclear fuels, nuclear power reactors, and commercial applications satellites. IBM holds it in computers. The oil-producing countries have begun to exert it, and it is feared that other producers of primary commodities will learn from their experience. This means that policies designed to 'optimise' the UK's (or any other country's) contribution to new technology within the framework of a perfect world market would be wrong, even if it could be translated into operational terms. Hedging against monopoly risks must inevitably be a part of government policy, and it is likely to cost more than an 'optimum' policy.

But do these considerations justify the present pattern of government support, amongst sectors and amongst types of R&D activity? Referring to Table 2, does the development of civil transport aircraft in Europe deserve such strong support, given that the possibilities of abuse of US monopoly power would be very limited, and that for Western Europe's internal needs there exists the viable alternative of fast trains?

Vol.
197

Furthermore, does the existence of monopoly power necessarily justify a national policy aiming at complete technological autarky? Given that no one Western European country commands more than about 15% of the world's scientific and technical resources, such a policy is likely to be expensive and (in the light of French experience in nuclear energy, and British experience in aircraft) unsuccessful. Are there possibilities of intermediate policies between complete dependence and complete autarky: for example, the maintenance of a minimum technological capability 'on the shelf' as an insurance policy, as Western European countries seem to be doing in space technology? Assuming that a national competence in computer technology is a desirable objective of policy, is the only way to achieve this the provision of government subsidies to try and maintain national firms competing against IBM, as the British, French and West German Governments are doing? There is little public discussion of these questions, and in the literature they appear to have been discussed only by Zysman [51] in relation to Western Europe, although there is a more substantial body of literature in relation to problems of technical dependence in the less developed countries [8].

Vol. 4
1976

10. CONCLUSIONS

It will be clear from the above discussion that our knowledge of the nature and extent of imperfections in the processes of industrial innovation to which governments should direct their attention, as well as of the comparative efficacy of the range of policy instruments which governments do or could apply to overcome them, is still limited. Nonetheless, some tentative policy conclusions do emerge.

In particular, it is possible that governments allocate too many R&D resources to the high technologies, because of an exaggerated concern for problems of security of supply, and to commercial development activities because of a mistaken belief that certain industries cannot find the resources for potentially profitable, larger-scale innovative investments. In Western European countries, this form of government involvement has grown very quickly over the past twenty years. This author has argued elsewhere [38] that it has been stimulated in large part by the strong competitive position in international markets of US firms in the high technologies, where large-scale Federal expenditures on defense, space and nuclear reseach have acted as an effective subsidy, as well as a drain on other sectors of US industry. At the same time, it can be argued that too few government resources have been provided to scientific, technical and managerial services for firms in traditional industries, and to long-term applied research outside the high technology sectors.

The data in Tables 1 and 2 show in what sense the UK and other governments were significant culprits in 1976. Table 1 shows that, when compared with industry's own expenditures on R&D, those of the UK Government on

civilian, industry-related R&D were considerably less than that of the Governments of Canada and France, somewhat less than those of the USA, and about the same level as those of Belgium, FR Germany, Italy and Norway. But they were considerably more than those of the Dutch, Japanese, Swedish and (although we do not have strictly comparable data) Swiss Governments. In these latter countries, industry's own expenditures on R&D were—and still are —comparatively high.

Table 1 also shows the amount of government R&D resources devoted to sectors other than the high technologies was small in all countries for which data are available. Table 2 shows that, like the UK Government, those of other EEC countries and the USA devoted most R&D resources to the high technology sectors, although there are significant differences in emphasis. In the UK, aircraft got the lion's share; in continental Europe, nuclear energy got the highest priority; and in the USA, most of the money went on space.

Unfortunately, we do not have data for countries other than the UK on the relative importance of government expenditures on commercial development as compared to exploratory, longer-term R&D. However, an examination of the distribution of government R&D expenditures to industry, government laboratories and the universities suggests a stronger emphasis on commercial development expenditures in France and the UK than in other countries [40].

In addition, it is unfortunate that we do not have a detailed breakdown of R&D expenditures in Eastern European countries. All that one can say is that the USSR shares with France and the UK the doubtful distinction of being ahead in the world race in supersonic air transport and in the development of fast breeder reactors. And, as in the UK, there are difficulties in the USSR in getting the results of R&D translated into production [1].

Trade and productivity data suggest that it is this feature that distinguishes the UK most strongly from other Western countries. Given our present state of knowledge, it is not reasonable to assume that poor performance in innovation and diffusion have played a more important role than other factors encompassed in the economist's notion of 'technical progress'. Nonetheless, there are a number of reasons why policies for innovation and diffusion may be particularly important. Two of them have been touched on already: the opportunity cost of the UK Government's continuing, large-scale commitment to aircraft, nuclear and defence R&D since the end of World War Two; and the lack of self-reinforcing relationships amongst innovation, diffusion and growth. In addition, there is the absolute and precipitous decline of UK industry's own R&D expenditures since 1970 [10]. This could be of considerable long-term significance, if one believes that the balance of payments is the greatest constraint on UK growth, and that engineering exports of high technical content are critical to the balance of payments.

What is the role of the economist in relation to government policies for innovation? Within the government structure, the economist's role could be

Vol. 197

described as unexciting, necessary and negative. Unexciting because the messy field of the economics of innovation is not yet one where one can make practical use of the latest developments in the economist's theory and manipulative technique. Necessary because some very large sums of public money appear to be spent without some rather simple and obvious economic questions ever being asked. Negative because the answers to these questions may often lead to a 'no' to requests for funds for commercial expenditures. Even in the case of longer-term projects, the answer should never be an irreversible, unqualified 'yes': it is very unlikely indeed that economists can do what no-one else can do, namely, predict with any degree of accuracy the outcome of R&D projects. Continuous review and appraisal will always be necessary.

Outside the government machinery, there is a considerable field for study. For this writer, the most important, both for theory and policy, is a better understanding of the decision rules which firms apply to industrial innovation (or lack of it), and the way in which these decision rules are affected by factors internal and external to the firm.[6] This is no easy task, since the observation by Nelson and Winter [34] that differences between industries in their technological and institutional characteristics may have important influences on the sources, nature and incentives for innovation has been amply borne out by this paper. Industries differ in the degree to which they use the public, R&D infrastructure, and are capable of using or pushing forward the scientific and technical state-of-the-art. They differ in the degree of technical sophistication and accountability of users. They differ in the costs of innovation, the ease of imitation, and the possibilities for rapid expansion. And they differ in the degree to which they are of strategic importance to the functioning of the modern industrial system, and are influenced by defence technology.

Vol. 4
1976

This suggests that some models encompassing the above factors (as well as others that might be important) should be built, and tested to see if they can explain differences in the dynamics of innovation amongst industrial sectors. It also suggests that empirical studies should be industry-specific. For UK public policy, one strong candidate would be an international comparison of innovation and the factors affecting it in the engineering industries.

In addition, there is ample scope for study of the effectiveness of the various policies and policy instruments used by governments in attempting to promote industrial innovation. There remain a large number of expensive, government-financed R&D programmes which have not yet been subjected to detailed, critical, academic scrutiny.[7] Work similar to Gardner's [17] remains to be done for other government programmes for financing commercial development. More needs to be understood about the workings of the government-financed

[6] In other words, why innovate (or not), and what innovations? Most recent empirical work has not concentrated on these strategic questions, but on the tactical (although important) question—how to innovate?

[7] To their credit, journalists have been energetic and outspoken [9, 21].

R&D infrastructure. Two main difficulties lie in the way of such research: the difficulties of obtaining accurate and significant information, and (particularly in the UK) the all-enveloping cloak of government secrecy. Examples from past studies show that both can be overcome.

Finally, in discussing and analysing government policies towards innovation, it would be better to start from 'political economy' rather than 'economics'. Political choice is important when deciding the degree of priority to be given to, say, the environment or to security of supply.[8] Furthermore, even within the ambit of specific policy objectives, choices about R&D and innovation are bound to remain political. As Freeman [16] has pointed out, R&D and innovation are always uncertain activities where no-one can be sure of the outcome. As a consequence, discussions and priorities, whether in industrial firms or in government, inevitably involve controversy, debate and advocacy, and political pressures. Although, it might be prevailing bureaucratic wisdom that 'R&D policy', 'innovation policy' or 'technology policy' should be subcomponents of 'industrial policy' or some other policy, it is a sociological and political fact that there are often coherent bodies of scientists and engineers, working in or outside government and acting as lobbies for government support of particular R&D programmes.

This is a legitimate and, given the uncertainties mentioned above, inevitable social activity, from which the public often benefits. But, as Eads and Nelson have argued in the case of US Government support of supersonic air transport and fast breeder reactors, and Gowing has shown in the case of UK Government support of nuclear energy, such lobbies can, in the absence of public debate and accountability, lead to the misallocation of resources [12, 20]. In analysing or advocating any government policy towards innovation, these factors must be kept in mind. Their existence suggests that, as in innovation within industrial firms, the economist's contribution to understanding will be all the greater to the extent that he recognises the potential contribution to be made by other disciplines and, in particular by the political scientists, the sociologist, the natural scientist and the engineer.

REFERENCES

1. Amann R (1970) The Soviet research and development tradition: the pressures of academic tradition and rapid industrialisation. *Minerva* **8**(2), 217–241.
2. Battelle (Columbus) (1973) *Interactions of Science and Technology in the Innovation Process: Some Case Studies.* National Science Foundation, Washington.
3. Burn D (1967) *The Politcal Economy of Nuclear Energy.* Institute of Economic Affairs, London.
4. Burns T and Stalker G (1961) *The Management of Innovation.* Tavistock.

[8] See, for example, the influence of President de Gaulle's concern about French independence on French innovation policy in the 1960's [19, 37].

5. BYATT I and COHEN A (1969) *An Attempt to Quantify the Economic Benefits of Scientific Research*. Science Policy Studies, No. 4, HMSO.
6. CAIRNCROSS A (1972) Government and innovation, in (Ed. WORSWICK G). *Uses of Economics*, Blackwell, Oxford.
7. CLARKE R (1973) Mintech in retrospect. *Omega* **1**(1), 25–38 and **1**(2), 137–163.
8. COOPER G (Ed) (1973) *Science, Technology and Development: The Political Economy of Technical Advance in Underdeveloped Countries*. Cass.
9. COSTELLO J and HUGHES T (1971) *The Battle for Concorde*. Compton Press.
10. Department of Industry (1974) Industrial expenditure on scientific research and development: provisional results of the 1972 survey. *Trade and Industry*, 5th September.
11. DÖRFER I (1973) *System 37 Viggen: Arms, Technology and the Domestication of Glory*. Scandinavian University Books.
12. EADS G and NELSON R (1971) Government support of advanced civilian technology. *Publ. Policy* **19**(3), 405–428.
13. FABIAN Y, YOUNG A *et al.* (1975) *Patterns of Resources devoted to Research and Experimental Development*. OECD, Paris.
14. FALK C (1974) Dynamics and forcasting of R&D Funding. *Technological Forecasting and Social Change*. **6**, 171–189.
15. FORES M (1972) Engineering and the British economic problem. *Quest*.
16. FREEMAN C (1974) *The Economics of Industrial Innovation*. Penguin.
17. GARDNER N (1976) Economics of launching aid, in Whiting, A. (Ed) *The Economics of Industrial Subsidies*, HMSO, London.
18. GIBBONS M and JOHNSTON R (1974) The roles of science in technological innovation. *Res. Policy* **3**(3), 220–242.
19. GILPIN R (1968) *France in the Age of the Scientific State*. Princeton University Press.
20. GOWING M (1974) *Independence and deterrence: Britain and atomic energy*, 1945–1952. Macmillan, New York.
21. GRAY R (1970) *Rolls on the Rocks*. Compton Press.
22. HARTLEY K (1974) *A Market for Aircraft*. Institute for Economic Affairs, Hobart Paper 57, London.
23. Illinois Institute of Technology (1969) *Technology in Retrospect and Critical Events in Science (TRACES)*. National Science Foundation, Washington.
24. JEWKES J, SAWERS D and STILLERMAN R (1969) *The Sources of Invention*. Macmillan New York.
25. JEWKES J (1972) *Government and High Technology*. Institute of Economic Affairs, Occasional Paper No. 37, London.
26. JOHNSON P (1973) *Co-operative Research in Industry*. Martin Robinson.
27. LITTLE A Inc (1973) *Barriers to Innovation in Industry: Opportunities for Public Policy Changes*. Washington.
28. MCCUTCHEON R (1975) Technical change and social need: the case of high rise flats. *Res. Policy* **4**(3), 262–289.
29. MORAND J (1968) La recherche et la développement selon la dimension des entreprises. *Le Progrès Scientifique* (122), Paris.
30. MYERS S and MARQUIS D (1969) *Successful Industrial Innovations*. National Science Foundation, Washington.
31. NABSETH L and RAY G (1974) *The Diffusion of New Industrial Processes*. Cambridge University Press.
32. NELSON R (1959) The simple economics of basic research. *J. Pol. Econ.* 297–306.
33. NELSON R, PECK M and KALACHECK E (1967) *Technology, Economic Growth and Public Policy*. Brookings.
34. NELSON R and WINTER S (1974) Neo-classical vs. evolutionary theories of economic growth: critique and prospectus. *Econ. J.* **84**(336), 886–905.
35. OECD (1964) *The Residual Factor and Economic Growth*. Paris.
36. OECD (1966) *Government and Technical Innovation*. Paris.
37. PAPON P (1975) The state and technological competition in France. *Res. Policy* **4**(3), 214–244.

Vol. 4
1976

38. PAVITT K (1974) International technology and the US economy: is there a problem?, in *The Effects of International Technology Transfers on US Economy*. National Science Foundation, Washington.
39. PAVITT K (1976) The Choice of Targets and Instruments for Government Support of Scientific Research, in Whiting, A (Ed), *The Economics of Industrial Subsidies*, HMSO.
40. PAVITT K and WALKER W (1976) Government policies towards industrial innovation: a review. *Res. Policy* **5**(1), 11–97.
41. PECK M (1968) Science and technology, in *Britain's Economic Prospects* (Ed CAVES R), Brookings.
42. ROTHWELL R and TOWNSEND J (1973) The communication problem of small firms. *Res. Dev. Magmt* **3**(3), 151–153.
43. SAPOLSKY H (1972) *The Polaris System Development: Bureaucratic and Programmatic Success in Government*. Harvard University Press.
44. SCHERER F (1973) *Industrial Market Structure and Economic Performance*. Rand McNally.
45. SEILER R (1965) *Improving the Effectiveness of Research and Development*. McGraw-Hill, New York.
46. TAYLOR C and SILBERSTON Z (1973) *The Economic Effects of the Patent System*. Cambridge University Press.
47. VERNON R (1966) International investment and international trade in the production cycle. *Q. J. Econ.* **80**(2), 190–207.
48. VERNON R (Ed) (1974) *Big Business and the State*. Macmillan, New York.
49. WESTIN A (Ed) (1971) *Information Policy in a Democracy*. Harvard University Press.
50. WILLIAMS B (1967) *Technology, Investment and Growth*. Chapman Hall.
51. ZYSMAN J (1975) Between the market and the state: dilemmas of French policy for the electronics industry. *Res. Policy*.

ADDRESS FOR CORRESPONDENCE: *Mr K Pavitt, Science Policy Research Unit, University of Sussex, Mantell Building, Falmer, Brighton, Sussex BN1 9RF, UK*

O MEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Analysis in Support of Strategic Policy Making[1]

KJ RADFORD

(*Received January 1976*)

Policy making involves choice between alternatives in situations that are usually very complex. The process of choice in such situations is not such that a straightforward analytical procedure can be prescribed yielding a uniquely rational solution. However, decision problems arising in policy making situations can be investigated using methods based on recent developments in game and metagame theory. This methodology can form the basis of a rational decision structure of the type envisaged by Diesing that can be incorporated in the day-to-day work of most organizations.

Vol. 4
1976

INTRODUCTION

MANY OF the problems that must be dealt with in strategic policy making are concerned with choice between alternatives and in this respect they are similar to those encountered in other decision-making activities. However, choices between policies are among the most complex decision problems because they usually arise in situations having one or more of the following characteristics:

available information may be insufficient for the construction of a comprehensive model of the situation and gathering of more information may be costly and time-consuming;

the situation may be new, and past experience in similar situations not directly applicable;

multiple objectives may be involved and priorities to be observed between objectives may not be defined clearly;

it may not be possible to assess progress towards achieving objectives in terms of well-understood and measurable parameters;

[1] Presented at the Second International Research Conference on Operational Research, Stratford-on-Avon, England, 12–15 April 1976.

there may be uncertainty concerning future events and the actions of others;

there may be more than one party involved in the situation with power to influence choice between alternatives;

any one situation may be inter-related with other concurrent situations and choices made in one area may have repercussions in others.

It is often very difficult (and sometimes impossible) in such situations to use the quantitative methods of evaluation and comparison between alternatives that have been developed for use in more familiar decision situations. While there has been considerable and continuing success in applying these methods to problems of a repetitive nature where past experience can be used as a guide in later encounters, analytical methods have not yet been devised that take full account of the complexity of most modern planning and policy problems. Furthermore, those responsible for decisions are usually observed taking a non-analytical approach to such problems and sometimes adopting what Lindblom has called "strategies and dodges" [1] in arriving at a choice between policies.

These intuitive approaches to complex decision problems probably result from the sub-conscious abandonment of the desirable (but practically unachievable) goals of uniquely rational choice and attainment of an 'optimum solution'. They may represent a transferring of the concept of rationality away from the component phase of choice between alternatives and its application instead to the construction of the decision procedure itself. This transfer was proposed originally by Diesing, when he described a rational decision structure as one that ". . . yields adequate decisions for complex situations with some regularity" [2]. The question before us is whether such a structure can be found to treat the complex decision situations that arise in strategic policy making.

THE CONSTITUENTS OF A STRATEGIC POLICY-MAKING SITUATION

The constituents (or ingredients) of a strategic policy-making situation are (a) one or more participants who have the power to influence the choice of policies; (b) a range of policy choices open to the participants and (c) a number of factors embedded in the situation that affect the participants' preferences for individual policies.

The participants in a policy making situation are those parties who can exert some influence over the outcome either by choosing and implementing a policy themselves or by interacting with those that do. The degree of participation is a function of the *potential* to exert influence rather than of the degree of open

involvement. Information available to participants is often incomplete and different participants usually have different sets of information in their possession at any time. However, two or more participants may have a common subset of information and may, in fact, exchange information in order to increase this common sub-set. A participant's perception of the situation in which he is involved depends upon the information available to him and the nature of the motivations, spheres of competence, experience and judgement of the person or persons involved. Each participant's perceptions of the situation are normally subjective and incomplete, although there is a natural tendency for those who have studied a problem to assume that *they* have the 'facts' and what *they* perceive is the real and 'objective' situation.

Each participant is motivated to consider a range of courses of action that he might undertake in order to transform what he considers to be a less desirable present situation into a more desirable future one. The nature of this range of courses of action depends firstly on the participant's appreciation of the broad requirements of the situation. He might, for example, consider only incremental alterations in policies to be appropriate or then again that more fundamental changes were needed [3]. Following this initial choice of the type of measure required, policy makers usually undertake a more detailed study of the courses of action (or options) that are considered to be *feasible* in the situation as it is perceived. The feasibility of any contemplated course of action depends on a number of factors that can be summarized as follows:

Vol. 4
1976

constraints, of a human, physical, technological or economic nature that apply to the individuals, organizations and physical objects involved in the policy situation;

policies, guidelines, rules and precedents that may be construed by a participant as limiting the range of available options;

standards of behaviour expected of individuals and organizations that may make unacceptable options otherwise regarded as feasible;

relationships and agreements between participants in the particular situation under consideration and between those in that situation and others in related and concurrent situations;

means and instruments for implementation of an option if it were finally selected;

patterns of power and authority that may enhance or reduce the feasibility of implementation of any particular option.

The set of options that each participant considers to be feasible depends upon his perception of the above factors in the light of the information available to him. Different participants may consider a different set of options feasible.

Furthermore, one participant's perception of what a second considers to be his feasible set of options may be very different from that which the second participant has derived himself.

CHOICE BETWEEN FUTURE SCENARIOS

Having developed a set of feasible options, a participant can proceed to consider the future situations (scenarios) that would result if he implemented one (or a combination of more than one) of these options. A future scenario can be regarded as an outcome of the choice of the option or options and the work of implementation. Participants in the policy-making situation normally have preferences between the various possible future scenarios. These preferences derive ultimately from values held by the individual or group involved as a participant. They may also depend on estimates of the feasibility and desirability of the course or courses of action required to bring about the future situation.

If only a single participant is involved in the policy-making situation, the process of choice is simply that of determining which of the future scenarios is most preferred by that participant. This may be done by simple inspection and judgement, assisted as necessary by appropriate quantitative analysis. In more difficult cases, where no quantitative methods can be used and where multiple objectives are involved, sequential elimination techniques [4] and their many variations may be useful in establishing which of the future scenarios is most preferred and therefore which of the options or policies should be adopted.

When two or more participants have the potential of influencing the choice of policies, the process is not so simple. This is particularly the case if the various participants have different perceptions of the policy-making situation and different preferences for outcomes and for methods for achieving them. In such cases, the choice of policies and options by one participant is influenced by that of another and vice-versa. The interaction between the participants is such that no one can achieve his most preferred outcome if this is incompatible with the wishes of the others. Resolution of such situations cannot therefore be achieved in terms of optimum solutions for any one participant. It may be sought, however, in terms of cooperative arrangements between all of them. A feature of such arrangements is that each participant does not achieve his most preferred outcome, but each achieves more by entering into a negotiated settlement than he might if he does not cooperate.

One of the focal points around which settlements can often be negotiated is called an *equilibrium*. An equilibrium is defined in game theory as a situation that can be brought about by appropriate choices of options by the participants from which it is not advantageous for any of them to move, *provided the others do not*. A settlement at an equilibrium allows the participants to avoid the worst

of their possible outcomes at the expense of not achieving their best. It may provide them with a measure of stability that they find desirable in a competitive environment. There is, however, no guarantee that all participants will seek the stability of a settlement at an equilibrium in any situation. One or more may prefer to seek an advantage in some other form of resolution of the problem. Even if they do seek the stability provided by an equilibrium position, more than one such position may exist. Furthermore, because these different positions offer different relative advantages to the participants, it may be difficult for them to decide amongst themselves which equilibrium position should be chosen as the basis of an agreement.

THE STRATEGIC STRUCTURE OF THE SITUATION

It is possible to investigate the stability or otherwise of future scenarios by a method based on recent developments in game and metagame theory and called the *analysis of options* or sometimes *metagame analysis* [5, 6]. In using this method, a scenario is tested to determine whether any one of the participants (or any coalition of participants) can expect to gain by moving away from it; that is whether he would prefer another scenario derivable from it. If no participant or coalition sees an improvement in his position resulting from a move away from a particular scenario, that scenario is by definition, an equilibrium and may possibly be the basis for a stable agreement, contract or treaty. If one or more participants or coalitions can improve their position by moving away from a particular scenario, the method allows consideration of what sanctions might be employed by the others to counter this gain. If credible sanctions exist, they may have the effect of maintaining the scenario under consideration as an equilibrium. If no sanctions exist, the scenario is not an equilibrium and one or more of the participants or coalitions can expect to gain by moving away from it.

In a practical application of the method, a number of possible future scenarios are usually examined for stability in this manner. As the analysis progresses, understanding of the problem increases and the results provide an indication of the strategic structure of the problem. This is illustrated best by an example.

Let us suppose that an electric power commission is required to formulate a policy on the size of nuclear power plants that it needs to construct to meet the anticipated demand. Technical and economic factors favour the 'nuclear park' concept, in which costs can be lower due to sharing of certain facilities between a group of closely spaced reactor sites. This concept is unpopular with local residents, because they fear disruption of their way of life, expropriation of farm and recreation land and possibly dangerous pollution of the area. However, construction of a number of smaller facilities (considered to be economically

Vol. 4
1976

unjustifiable by the commission) would bring employment to the areas in which they were located and this policy is favoured by the local citizens' groups. A compromise policy would involve mid-sized facilities that would still provide local employment but at a lower level of disruption of the local scene and possible pollution danger.

The commission has publicly announced its intention to proceed on the basis of a policy of installing a large nuclear park, but it needs the support of the government to continue with this policy. A citizens' group has been organized in the area concerned with the objective of bringing pressure to bear on the government by picketing the proposed site and other similar measures. The participants and their options in this much simplified example can be listed as follows:

Participants	Options	Present situation	Proposed compromise
Power commission	Adopt policy of large site	1	0
	Adopt policy of small sites	0	0
	Settle for mid-size sites	0	1
Citizens' group	Picket proposed site	1	0
	Withdraw support from government	0	0
	Settle for mid-size sites	0	1
Government	Support commission in large site policy	1	0
	Support mid-size site policy	0	1

Scenarios are constructed by placing a 1 or 0 against an option to indicate that the particular option is exercised or not. A complete scenario is indicated by reading down the column. For example, the present situation is that the power commission is adopting a large site policy and not the other alternatives, the citizens' group is picketing the proposed site but not as yet withdrawing its support from the government or indicating willingness to settle for mid-size sites and the government is supporting the commission in its large site policy. The proposed compromise is indicated in a similar manner. The analysis allows consideration of whether the compromise would be stable for all participants and coalitions of participants. Lack of space in this paper requires that details of the analysis be omitted. Readers wishing to pursue these details can construct them by reference to similar analyses reported in the literature [7, 8].

The results of the analysis can be illustrated in a diagram that shows the strategic structure of the problem such as in Fig. 1. In this diagram, the compromise position is shown to the left, with three improvements for the coalition of the commission and government emanating from it. This indicates that in the analysis, that coalition prefers (given a unilateral choice) the three situations shown at the ends of the arrows from the compromise. However, once the

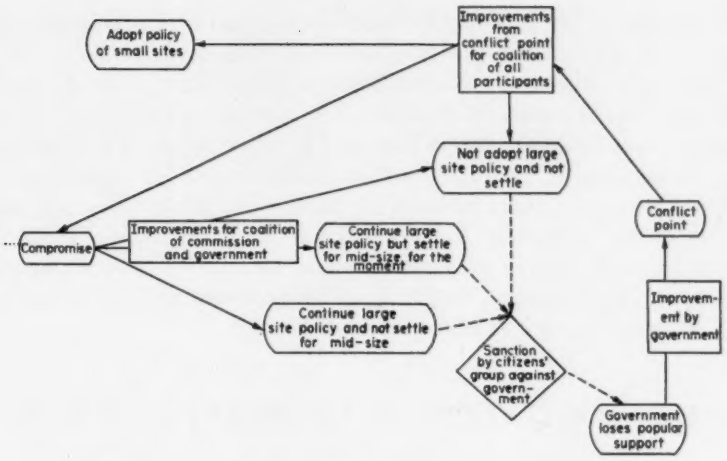

Adopt policy of small sites
Improvements from conflict point for coalition of all participants
Not adopt large site policy and not settle
Conflict point
Compromise
Improvements for coalition of commission and government
Continue large site policy but settle for mid-size, for the moment
Improvement by government
Continue large site policy and not settle for mid-size
Sanction by citizens' group against government
Government loses popular support

FIG. 1. *Strategic structure of the nuclear power policy problem.*

Vol. 4
1976

interaction of the citizens' group's preferences are considered, it is revealed in the analysis that a move by that participant to bring about the defeat of the government at a forthcoming election would constitute a credible sanction against these improvements. This would lead to an outcome labelled 'government loses popular support' at the bottom right of the diagram. From this situation, an immediate improvement can be seen for government in terms of breaking the coalition with the commission and bringing pressure to bear on that participant to modify its policies in order to retrieve some measure of popular support.

This leads to a situation known as the 'conflict point', which is a standoff position at which no participant is achieving his objectives. The commission believes it will lose government support if it continues its large site policy, the government fears loss of popular support if it does not pressure the commission into a mid-size or small site policy and the citizens' group stands to lose the benefits of construction in its local area if no decision is made. This conflict point is typical of the impasse that arises in many policy-making situations of this type. If no participant is willing to concede, the situation may remain at this point for some time. The analysis shows, however, three improvements from the conflict point that can be obtained by agreement (coalition) between all three participants. These improvements are (i) the adoption of a policy of small sites, preferred by the citizens' group (ii) the compromise, probably preferred by government after appraisal of the situation and (iii) the non-cooperative position of not adopting a large-site policy but not agreeing to the compromise, that might be preferred, at least for a while, by the commission. This last outcome is, however, shown in the diagram to be subject to the sanction

by the citizens' group. It is therefore part of a loop at the right hand side of the diagram that is unlikely to lead to a stable resolution of the problem.

Which of the other two outcomes shown in the diagram is actually adopted by the participants cannot be decided by the analysis. In practical situations, it would probably be determined in large part by the strengths of the preferences of the participants, the credibility of the sanctions available to them and the course of any negotiations between the participants aimed at mutual escape from the conflict point. The analysis does not therefore present a single 'best' solution to the problem. It provides instead an overview of the strategic moves available to the participants. Resolution of the problem requires a further stage that is concerned more with tactics than with strategy.

THE NEGOTIATION AND BARGAINING STAGE

The problem of deciding what tactics should be employed in such a situation is complicated by the fact that the participants have only limited amounts of information. Furthermore, different participants usually have different information and therefore may have different perceptions of the various aspects of the situation. Any two participants conducting an analysis of the type just described may therefore make different assumptions concerning:

(a) the number and identity of participants;
(b) the options available to these participants;
(c) the preferences of these participants for possible future scenarios; and
(d) the likelihood of possible future natural events and their effect on all the above factors.

The strategic structure of the policy situation as viewed by one participant may therefore be different from that perceived by another. These discrepancies can, of course, be corrected to some extent by communication between the participants and this is what is usually attempted in decision making groups in which all members subscribe to a common purpose or objective. However, in situations where some conflict of interests and objectives exists between the participants, complete communication of his perceptions may be to the disadvantage of any one of them. The resolution of the policy situation may then be attained only after a period of negotiation and bargaining, during which some amounts of information are passed between the participants. An important question to be resolved by each of the participants in such circumstances concerns the amount of information about his perceptions of the situation he should pass to the others and in what form should this information be released. In short, what tactics should he adopt in the negotiation and bargaining processes?

A detailed treatment of this subject is outside the scope of this paper. It can be noted, however, that it has been discussed by Schelling in his classic work *The Strategy of Conflict* [9], which can be considered (in the light of the arguments presented here) to be more concerned with tactics that might be adopted in a conflict situation than with the underlying strategic structure. These tactics range from the issuing of threats, warnings and promises to the communication of an irrevocable commitment. The information that is passed when employing one or more of these tactics may be true or false or somewhere in between, depending on the extent to which the communicator is bluffing. However, the objective in every case is to influence the other participants and to induce them into accepting a resolution that is desired by the participant making the communication.

THE PROCESS OF STRATEGIC POLICY MAKING

Strategic policy making is therefore seen as a continuing process that is closely linked to the more tactical activities of communication, negotiation and bargaining between the participants in the situation under consideration. In typical practical circumstances, the process can be envisaged as consisting of a continuing series of sessions with the time interval between each depending on the urgency of the problem under consideration (Fig. 2). In an average case, this time interval might be weeks or even months, although in cases of extreme urgency the discrete sessions might be much closer together. At times they might comprise an essentially continuous consideration of the situation as appears to have been the case for both major participants in the Cuban Missile Crisis [10].

The agenda for each of these sessions would consist of the following topics:

1. review of the strategic structure of the situation and of tactical measures agreed at the last session;
2. introduction of new information gathered since the last meeting on both natural events and the results of tactical actions;
3. reconsideration of the strategic structure of the situation in the light of this new information;
4. decisions concerning tactical actions to be taken at this time.

The procedure illustrated in Fig. 2 is essentially that normally followed by skilled managers and administrators faced with a complex policy problem [11]. For this reason it is likely to be readily understood by those involved when introduced into an organization as the basis for consideration of strategic policy situations. Furthermore, once the procedure is installed, it is unlikely to intrude into the actual work of formulating policy.

Vol. 4
1976

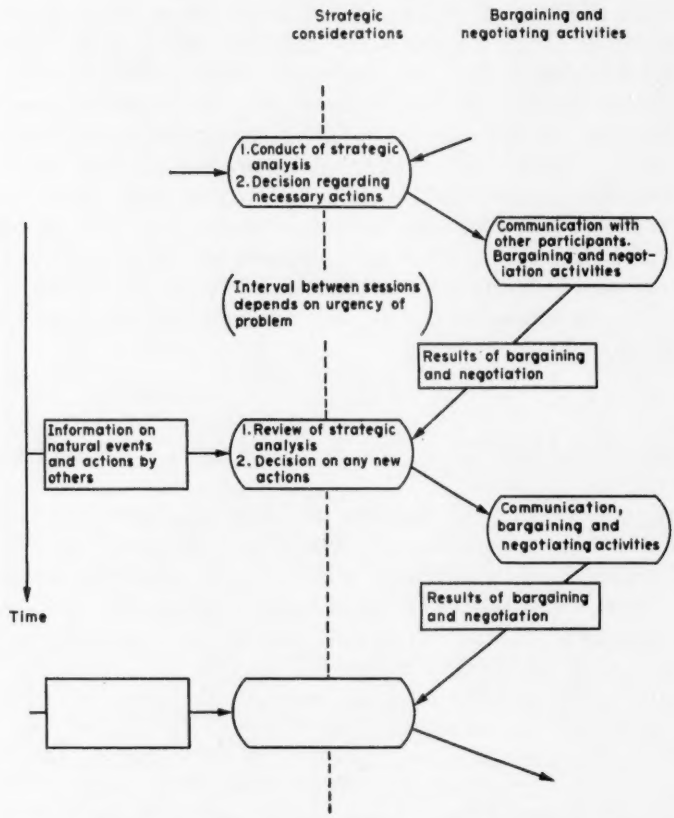

Strategic considerations
Bargaining and negotiating activities
1. Conduct of strategic analysis
2. Decision regarding necessary actions
Communication with other participants. Bargaining and negotiation activities
Interval between sessions depends on urgency of problem
Results of bargaining and negotiation
Information on natural events and actions by others
1. Review of strategic analysis
2. Decision on any new actions
Communication, bargaining and negotiating activities
Results of bargaining and negotiation
Time

FIG. 2. *Diagrammatic representation of the strategic policy making process.*

SUMMARY

Strategic policy problems are not usually amenable to treatment by the quantitative methods developed to deal with the more repetitive decision situations that arise in routine operations and administration. Intuitive approaches to such policy problems do not seek a uniquely rational optimum solution. Instead, the concept of rationality is applied to the selection of a decision structure within which the problem can be considered and which provides an adequate means of resolution.

Recent developments in game theory have provided a basis for analysis of these problems in terms of the participants and coalitions that can influence the outcome of a policy situation, the options or courses of action open to them

and their preferences for future scenarios and outcomes. By consideration of these factors, the analysis allows the strategic structure of a policy situation to be determined as a preliminary to decisions regarding tactical moves that can be made by the participants in an effort to resolve the problem.

The process of strategic policy making is seen as a continuing process closely linked to the more tactical activities of communication, negotiation and bargaining. A procedure for strategic policy making can be envisaged based on the form of analysis described that can be incorporated unobtrusively into the day-to-day work of an organization.

REFERENCES

1. LINDBLOM CE (1968) *The Policy Making Process*, p. 24. Prentice-Hall, Englewood Cliffs.
2. DIESING PAUL (1962) *Reason in Society*, pp. 171–198. University of Illinois Press.
3. ETZIONI A (1968) *The Active Society*, pp. 282–312. McMillan, New York.
4. MCCRIMMON KR (1973) An overview of multiple objective decision making in *Multiple Criteria Decision Making*, (Eds. COCHRANE JL and ZELENY) pp. 18–44. University of South Carolina Press.
5. HOWARD N (1971) *Paradoxes of Rationality*, pp. 172–146. MIT Press, New York.
6. RADFORD KJ (1975) *Managerial Decision Making*, pp. 148–172. Reston, New York.
7. HOWARD NIGEL (1975) The analysis of options in business problems. *INFOR* The Canadian Journal of Operational Research and Information Processing, **13**(1), 48–67.
8. BAIN H, HOWARD N and SAATY TL (1971) Using the analysis of options technique to analyze a community conflict. *Journal of Conflict Resolution*, XV (2), 133–144.
9. SCHELLING TC (1963) *The Strategy of Conflict*. Oxford University Press, London.
10. ALLISON GRAHAM T (1969) Conceptual models of the Cuban missile crisis. *American Political Science Review*, LXIII, 689–718.
11. RADFORD KJ *Dealiing with Complex Decision Problems*, Reston, New York. To be published.

ADDRESS FOR CORRESPONDENCE: *Professor KJ Radford, Department of Administrative Studies, Atkinson College, York University, Downsview, Ontario, Canada.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Dynamic Strategies for Branch and Bound

FRED GLOVER

LEE TANGEDAHL

(*Received July 1965; in revised form June 1976*)

Some straightforward but unconventional strategies are proposed for implementing branch and bound methods. These include 'shrinking' the branch and bound tree and instituting 'branch reversals' by reference to the notion of the relative influence of particular branches in the current solution. An attractive feature of these strategies is their ease of implementation, and the fact that they free the solution process from its customary dependency on early branches created on the basis of inadequate information.

Vol. 4
1976

A GOOD many proposals for improved branch and bound methods develop rather intricate types of penalty calculations, problem relaxations, or 'multi-fork' branching alternatives (see, e.g. [1–8]). This note instead proposes a variety of straightforward strategies that are easily implemented, but that involve manipulating branches in non-standard ways. The following reflections and observations are offered informally as a possible spur to investigate by-passed alternatives.

One of the more vexing aspects of branch and bound is the fact that branching decisions must be based on very limited information about the likelihood that a particular branch will lead to an optimal solution. This limitation is especially pronounced at early stages of the branch and bound tree. At such stages, the small number of imposed branches does not yield sufficiently accurate bound calculations or other measures of branch desirability to provide very reliable estimates about which branch should preferably be imposed next. Yet the choices made at early stages can have a dominating influence on the efficiency of the branch and bound process. Even 'reasonable' choices can be extremely poor if they are not sufficiently influential to reduce the alternatives for other variables. Unless a current branch has the power to inhibit the range of remaining alternatives, the branch and bound process can degenerate into the disastrous approximation to total enumeration sometimes documented.

This research was partly supported by the Office of Naval Research, Contract N00014-76-C00383 with Decision Analysis and Research Institute, Inc., Boulder, Colorado, USA.

From this point of view, it seems worthwhile to consider a branching technique that has the ability to rid itself of certain types of uninfluential branches on the basis of more reliable information available at later stages. A chief component of this technique is to shrink the branch and bound tree by eliminating earlier branches that have been rendered 'conditionally superfluous' by subsequent decisions.

In particular, suppose that an integer variable x_j has been subjected to a branching constraint such as $x_j \leq k$ or $x_j \geq k$, where k is an integer. This branching constraint will be called 'currently uninfluential' if the updated LP objective-function coefficient for its slack variable ($s_j = k - x_j$ or $s_j = x_j - k$) is 0. To understand the motivation for this terminology, note that if x_j has been subjected to the restriction $x_j \leq 5$, and x_j receives a value of $4\frac{1}{8}$ in the current LP solution, then the slack variable $s_j = 5 - x_j$ is positive and hence basic. This implies the LP objective-function coefficient for s_j is 0. Consequently, the branch $x_j \leq 5$ qualifies as currently uninfluential—which is appropriate, because the branch does not actively restrict the value of x_j in the current solution. More generally, the stipulation that the objective-function coefficient for s_j is 0 is motivated by the fact that this implies the branching constraint does not affect the optimality of the current LP solution.

Regardless of the sequence in which branches are historically imposed, it is of course always possible to defer the decision about the sequence in which they are viewed as being imposed, for the purpose of implementing various tree search rules (such as the LIFO and 'best bound' rules). The only restriction to this deferred sequencing decision is that branches that historically precede a compulsory branch (that is, one whose alternative has been eliminated by examination or fathoming) cannot be re-sequenced to follow this branch. (This follows simply from the fact that compulsory branches rely on historically antecedent branches for their compulsory status.) The preceding notion of a currently uninfluential variable suggests the strategy not merely of re-sequencing historical decisions, but of actually discarding some of them, thereby collapsing the branch and bound tree.

For instance, it characteristically happens that some variables will automatically receive integer variables (e.g. 0) as a consequence of imposing branch restrictions on other variables. Yet from lack of reliable local information, variables that may be 'dependent' on others cannot be distinguished from their companions, and thus may be incorporated into the branching decisions. The ability to identify and release branches that subsequently are discovered to be uninfluential in the indicated sense can therefore be useful. The only restriction that must be observed is that a branch cannot be discarded if it precedes a compulsory branch—again for the obvious reason that the compulsory branch may depend on the antecedent branch for its compulsory status.

The same type of notion that prompts this tree shrinking strategy also suggests the possibility of ranking branches in terms of their current influence, as

Vol.
197

measured by the magnitude of updated objective-function coefficients for slack variables associated with the branches. Such a ranking can be used to determine the sequential ordering of the branches in the branch and bound tree. For example, by analogy with the situation in which currently uninfluential branches may be discarded, more influential branches may be placed earlier in the tree and less influential branches later. (We shall call this a 'resequencing strategy.') This raises interesting issues, including a rationale for reversing as well as discarding branches. We briefly summarize these as follows.

(1) Since rankings will change depending on the current LP solution, the point at which a ranking imparts a sequence position to a given branch can make a difference in the structure of the branch and bound tree. This suggests keeping a history of influence measures for branches and biasing current influences in the direction of a weighted composite of historical influences.

(2) Influence measures communicate two competitive types of information. While it seems natural from one standpoint to place the more influential branches earlier in the tree, it seems equally natural from another standpoint to regard these branches as more likely to represent poor branching alternatives. In particular, the larger the objective-function coefficient attached to a given slack variable, the more 'expensive' the associated branch appears to be (since reversing the branch by allowing the slack to equal -1 will locally improve the objective function by the value of the objective-function coefficient).

These competitive information aspects can be made the basis of a 'branch reversal' strategy. For example, one may re-evaluate a seemingly expensive branch by setting its slack to -1 and re-optimizing with the dual method. Or more simply, one can use a single dual pivot update of the objective-function value. If setting the slack to -1 yields no infeasibility (or an infeasibility that is easily removed while maintaining an improved objective function value) then a branch reversal seems warranted. A reversal antecedent to a compulsory branch would destroy the compulsory status, but should at least be considered as a trial possibility upon reaching an integer feasible solution—since if such a reversal yields no infeasibilities it may immediately provide an improved candidate for an optimal mixed IP solution.

(3) The evaluation of branch reversal possibilities allows a refinement in the choice of a sequence for the current branches. Specifically, maintaining the view that influential branches should appear earlier in the sequence, it seems just as important for branches with relatively attractive alternatives to appear later, and an intelligent sequencing strategy will attempt to exploit both of these notions. Historical information concerning prior evaluation of branch alternatives can reasonably be used to supplement the sequencing decision.

All of these ideas can be implemented with relatively minor changes in current branch and bound procedures. They are proposed in this spirit, as simple

Vol. 4
1976

strategic notions whose time for study seems ripe—and that can be readily pursued by convenient modification of existing systems.

An example

To illustrate the operations of shrinking and resequencing, and to provide a clearer understanding of how they can affect the solution process, consider the following problem:

$$\begin{aligned} \text{Minimize } x_0 = \quad & 8x_1 + 8x_2 - 4x_3 \\ & 3x_1 \qquad - 2x_3 \geq 2 \\ & 2x_x - 2x_2 \qquad \geq 1 \\ & -8x_1 + 20x_2 \qquad \geq -1 \\ & x_1, x_2, x_3 \geq 0 \text{ and integer.} \end{aligned}$$

The following sequence of branch and bound steps, for one set of branching rules (whose form is unimportant for illustrating the shrinking and resequencing operations), identifies the values of the variables at each LP optimization by the vector $x = (x_1, x_2, x_3)$.

Step 1: Initial LP solution: $x = (0{\cdot}75, 0{\cdot}25, 0{\cdot}13)$
Branch: $x_1 \geq 1$.

Step 2: LP reoptimization: $x = (1, 0{\cdot}35, 0{\cdot}50)$
Branch: $x_3 \geq 1$.

Step 3: LP reoptimization: $x = (1{\cdot}33, 0{\cdot}48, 1)$
Shrink: drop $x_1 \geq 1$ Branch: $x_1 \geq 2$.

Step 4: LP reoptimization: $x = (2, 0{\cdot}75, 2)$
Shrink: Drop $x_3 \geq 1$ Branch $x_2 \geq 1$.

Step 5: LP reoptimization: $x = (2, 1, 2)$
Feasible integer solution: $x_0 = 16$
Resequence: current branch sequence $x_1 \geq 2, x_2 \geq 1$
new branch sequence $x_2 \geq 1, x_1 \geq 2$
Backtrack: impose $x_1 \leq 1$.

Step 6: LP reoptimization: no feasible solution
Backtrack: impose $x_2 \leq 0$.

Step 7: LP reoptimization: no feasible solution
Problem solution complete (no branches left).

In the preceding steps, note that the opportunity to shrink the tree occurred twice, the opportunity to resequence at backtracking occurred once, and these operations were implemented in each case. (In addition to the considerations

previously discussed, another useful criterion for resequencing, used here, is to position branches earlier in the tree for those variables whose values have covered the smallest range during the LP reoptimizations.) Using the same branch rules, the solution of the foregoing problem requires eleven steps (LP reoptimizations) if either the option to shrink or the option to resequence is bypassed, and requires fifteen steps if both options are bypassed.

Preliminary computational experience

To obtain a preliminary indication of the computational promise of dynamic manipulation strategies in branch and bound, we have programmed and tested the method on eight small problems, ranging from 17 to 25 integer variables and 9 to 21 constraints. In spite of their size, these problems have presented considerable difficulty to other solution methods (see [9, 10]). In fact, one of those problems, from DuPont, has more than 8 years of history as a tough problem, and the latest attempt to solve it with a commercial integer programming system required over 20 min on the IBM 360/65.

Our test results were extremely encouraging. The application of the shrinking and resequencing strategies resulted in cutting solution times for these problems by factors of approximately two to seven (21/12 sec and 71/11 sec) in the worst and best cases. The 'difficult' DuPont problem was solved in only 20 sec, using the CDC 6400. The question naturally arises whether the improvements from the shrinking and resequencing strategies were available only because the version of the B&B routine without them was poor. In fact, the opposite is the case. The B&B procedure in the absence of these strategies proved substantially more effective for solving the test problems than the methods cited in [9, 10]. This procedure succeeded in solving the DuPont problem, for example, in 87 sec.

More extensive tests on wider ranges of problems, employing branch reversal strategies as well as shrinking and resequencing strategies, are currently under way.

Vol. 4
1976

REFERENCES

1. ARMSTRONG RD and SINHA P (1974) Improved penalty calculations for a mixed integer branch-and-bound algorithm. *Mathl Prog.* **6**, 212–213.
2. BALAS E (1972) Integer programming and convex analysis: intersection cuts from outer polars. *Mathl Progr.* **2**(3),
3. BREU R and BURDET CA (1974) A subadditive approach to the group problem programming. *Mathl Progr.* **2**, 1–50.
4. FISHER ML, NORTHRUP WD and SHAPIRO JF (1974) Using duality to solve discrete optimization problems: theory and computational experience. Working Paper OR 030-74, Operations Research Center, MIT.
5. GEOFFRION AM (1974) Langrangean relaxation for integer programming. *Mathl progr.* **2**, 83–114.

6. GLOVER F (1973) Polyhydral annexation in mixed integer programming. MSRS 73-9, University of Colorado.
7. JEROSLOW RG (1974) Relaxations of integer programs. Management Science Research Report No. 347, Carnegie–Mellon University.
8. TOMLIN JA (1971) An improved branch and bound method for integer programming. *Ops Res.* **19,** 1070–1075.
9. WHITE CH (1976) Discrete blending (trouble in small packages). Paper presented at the Joint National ORSA/TIMS Meetings.
10. WOOLSEY RED (1973) Difficult integer programming problems. Colorado School Mines, Golden, working paper.

ADDRESS FOR CORRESPONDENCE: *Professor Fred Glover, Business Research Division, Graduate School of Business Administration, University of Colorado, Boulder, Colorado* 80302, *USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Information Processing in Managerial Decision Making: A Preliminary Study

JOSEPH G SAN MIGUEL

(*Received October 1975; in revised form June 1976*)

This paper reports the preliminary findings of an exploratory investigation into information processing in a managerial context. The vehicle for testing several research objectives was a laboratory decision-making experiment in which subjects purchased information items and evaluated a set of organizational goals. The methodology included multidimensional statistical methods that hold promise for dealing with the complexity of problems in implementing managerial decision models. Within a specific information–decision scenario, the statistical evidence favorably supports the existence of individual differences in information perception and information selection, and the use of goal preference hierarchies as predictors of information preferences.

Vol. 4
1976

THE ROLE of the information systems designer is that of satisfying users' wants and preferences for information for making decisions. This user emphasis extends to all areas of corporate decision making, such as production, marketing, research and development, finance, and personnel. The research reported here sheds light on two basic issues concerning information processing by decision-makers. First, the manner in which different information stimuli are perceived to have value in decision-making by users. Second, if different individuals or groups of individuals react differently to the same information stimuli, then an argument can be made for individual differences in information processing. The diversity of users' background, objectives and psychological characteristics may lead to different wants and preferences for information.

Since the area of information processing consists of an emerging set of inter-disciplinary tools, it is not possible to cover all of the issues. In view of the limited amount of published research, this exploratory paper is primarily concerned with methodological issues of data analysis.

The specific objectives of this research were:

1. to analyze the *similarities* of information items using a multidimensional approach;
2. to analyze the *preferences* for information items using a multidimensional approach;

3. to examine the efficacy of 'natural' groups of decision-makers for prediction of information selection.

METHODOLOGY

Briefly, a set of data was obtained from 87 undergraduate business students regarding similarity judgments on and preferences for items of information, and weightings of a set of goals.[1] The subjects were given a specific decision scenario concerning the role of the decider in a firm faced with uncertainty in the awarding of a contract. Given the contract uncertainty, the decider had to select one of the following alternatives: maintain the present level of operations; close down operations half way; or close down operations completely.

The specific steps of the laboratory experiment were:

1. A weighting assignment to three organization goals (employee relations, liquidity, and profits) using the Churchman–Ackoff technique [1]. This defined the *goal structure* for each individual.
2. Judgments concerning the *similarity* (or *proximities*) of a set of fifteen information items using bi-polar adjective pairs.
3. Purchase of information items *prior to decisions* subject to a budget constraint. The frequencies of purchase were measures of the *preferences* for the information items.

Subsequent to the experiment the goal weightings were normalized. The subject was unaware of this procedure. The bi-polar adjective pairs consisted of good–bad, strong–weak, important–unimportant, valuable–worthless, and useful–useless. The derived similarity scores were computed by summing the scores (ranging from 1 to 7) on the adjective scales. The maximum purchase frequency for any particular item of information was five, the number of opportunities to purchase the item.

1. *Analysis of proximities*

The KYST computer program is used to analyze the similarities (proximities) between information items [5, 8]. Distances between the data values (DIST) are used to cluster *similar* data values in *n*-dimensional space. The objective of the KYST algorithm was to find the best 'fit' for the fifteen points in a three dimensional space. The index of 'fit' used in most multidimensional scaling algorithms is Kruskal's stress measure [4], which is similar to an *F*-statistic.

The final configuration values for three-dimensional space are listed in Table

[1] A detailed description of the experiment and results may be obtained by writing the author.

Vol. 197

1, where the information items can be rank ordered on each dimension. In this case the dispersion of values is highest on Dimension 1 and lowest on Dimension 3.

TABLE 1. THREE-DIMENSIONAL SCALING OF PROXIMITIES OF INFORMATION ITEMS

Information item	Dimensional values in final configuration		
	1	2	3
Lines of credit available	1·64	0·26	—0.21
Predicted cash flows	1·59	0·25	−0·18
Cost, volume, profit behavior	1·25	0·14	0·08
Long-term lease commitments	0·83	−0·04	0·32
Return on capital employed	0·70	−0·11	0·33
Age of plant and equipment	0·04	−0·30	0·01
Strength of employees' union	0·01	−0·30	0·01
Funds invested in research and development	−0·25	−0·29	−0·11
Union contract commitments for continued payments	−0·30	−0·27	−0·11
Number of successful contract bids in the past	−0·36	−0·23	−0·14
Previous year's profits	−0·60	−0·10	−0·12
Market value per share of common stock	−1·06	0·21	−0.00
Net realizable value of inventory	−1·06	0·22	−0·00
Firm's role in community's economy	−1·16	0·26	0·04
Earnings per share	−1·28	0·31	0·09

Stress = 0·003

2. *Analysis of information preferences*

To establish the underlying dimensionality of information purchasing (preferences) behavior and to rank each subject's individual behavior along each dimension, the purchase frequencies were submitted to factor analytic techniques [3], which use the degree of correlation between items. A matrix of product-moment correlations between all pairs of information purchase frequencies was subjected to factor extraction by the principal axis method to produce a 15 × 3 matrix of *factor loadings*, which is given in Table 2. The information items with loadings of 0·5 or higher aid in identifying the common theme for each factor. For example, the theme of Dimension 1 relates to external financial information of an operating results character. The information items high on Dimension 2 portray a concern for societal-employee welfare type of information. Dimension 3 is directly concerned with projection of cash flows since this is the only item of information with a very high factor loading.

3. *Analysis of goal preferences*

To classify the subjects into 'natural' clusters or groups based on their manifest goal weightings on the Churchman–Ackoff procedure, the hierarchical grouping technique developed by Ward was employed [7]. The optimal grouping maximizes the average between-group distance while simultaneously minimizing

Vol. 4
1976

TABLE 2. FACTOR ANALYSIS OF PURCHASE FREQUENCIES FOR INFORMATION ITEMS

	Mean value	Factor loading for dimensions I	II	III
Predicted cash flows	3·99	0·04	0·11	0·88*
Strength of employees' union	1·48	0·18	0·53*	0·19
Number of successful contract bids submitted in past	1·85	0·22	0·10	−0·04
Firm's role in community's economy	0·97	−0·08	0·54*	0·01
Return on capital employed	2·51	0·59*	−0·01	0·25
Cost, volume, profit behavior	4·14	0·12	0·27	0·28
Net realizable value of inventory	1·99	0·35	0·06	−0·09
Previous year's profits	1·30	0·52*	0·14	0·05
Commitments on long-term lease	2·97	0·23	0·26	0·05
Earnings per share	0·81	0·82*	0·01	0·07
Market value per common share	0·71	0·62*	0·15	0·01
Age of plant and equipment	2·04	0·14	0·44	0·08
Funds invested in research and development	1·47	0·20	0·22	0·34
Available lines of credit	3·23	−0·01	0·16	0·46
Union contract commitments for employee benefits	1·99	0·15	0·72*	0·02
Eigenvalue		3·14	1·42	1·00
Percentage of variance explained		46·9%	21·2%	14·8%

* Value exceeds 0·5.

the average within-group distance. In this regard the hierarchical grouping method is 'sensitive' to the distances between the set of goal weights. Hence, similar goal weightings will cluster in homogeneous groups. The results from the five-cluster stage of the hierarchical grouping are presented in Table 3 in terms of the standardized goal means for each of the five groupings.

The relationship between individual goal preferences and information preferences can have very specific implications for the information systems

TABLE 3. HIERARCHICAL GROUPING ANALYSIS OF GOAL WEIGHTS

	Hierarchical 5-Groups A	B	C	D	E
Group size	22	20	11	27	7
Group means (standardized):					
Employee relations (=E)	−1·01	0·11	1·90	0·19	−0·84
Liquidity (=L)	0·24	0·45	−0·48	−0·94	2·34
Profits (=P)	0·73	−0·51	−1·35	0·67	−1·33
Group descriptors:					
Most dominant goal	P	L	E	P	L
Next dominant goal	L	E	L	E	E
Least dominant goal	E	P	P	L	P

designer. To explore this relationship, the technique of *multiple* discriminant analysis was selected [2]. Analogous to one-way analysis of variance, this technique attempts to maximize the *F*-ratio of between-means variance to the within-group variance for each variable in a set of dependent variables operating together (information items in our case).[2] A Wilks' Lambda value is used to test the statistical significance of the *F*-ratio [6]. The strengths of the discriminating variables can be captured in 'discriminant functions' resembling the factors in factor analysis. The results (using the five groupings of the subject's goal weights) are shown in Table 4, where only the first discriminant function was found statistically significant.

TABLE 4. MULTIPLE DISCRIMINANT ANALYSIS OF INFORMATION ITEMS FOR THE FIVE GROUPS

	Discriminant functions			
	I	II	III	IV
Percent of variance accounted for:	63%	18%	12%	7%
Chi-square statistical test: $p =$	0·00+	0·34	0·62	0·81
Centroid coordinates:				
Group A (P, L, E)	3·47	2·27	1·55	−0·34
Group B (L, E, P)	2·58	2·73	1·07	−0·38
Group C (E, L, P)	1·90	2·45	1·81	−0·62
Group D (P, E, L)	2·58	1·78	2·21	−0·51
Group E (L, E, P)	3·41	2·48	1·26	−1·20

1. Discriminant functions extracted accounted for 100% of the variance.
2. Statistical test of the overall group discrimination:
   Wilks' Lambda = 0·325.
   F-ratio = 1·491; probability = 0·02.

SUMMARY AND CONCLUSIONS

The research reported in this paper sought to determine: (1) whether, for a specific decision scenario, different information items are perceived differently by users; (2) if different information items are preferred over others with some underlying structure; and (3) whether goal preferences can be used as predictors of information preferences. The methods described suggest means for examining these issues. Because of the 'state of the art' of the subject area under examination we view this study as speculative and exploratory. The hope is that this research will permit identification of further ideas and hypotheses for future testing.

[2] The use of discriminant analysis for prediction purposes is also akin to multiple regression analysis, except that more than one independent group is being examined in discriminant analysis.

Vol. 4
1976

REFERENCES

1. CHURCHMAN CW, ACKOFF RL and ARNOFF EL (1957) *Introduction to Operations Research.* John Wiley, New York.
2. COOLEY WW and LOHNES PR (1971) *Multivariate Data Analysis.* John Wiley, New York.
3. HARMON HH (1960) *Modern Factor Analysis.* University of Chicago Press, Chicago.
4. KRUSKAL JB (1964) Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika* **29,** 1–27.
5. KRUSKAL JB, YOUNG FW and SEERY JB (1973) *How to use KYST.* Bell Telephone Laboratories, Murray Hill, New Jersey.
6. VELDMAN DJ (1967) *Fortran Programming for the Behavioral Sciences.* Holt, Rinehart & Winston, New York.
7. WARD JH (1963) Hierarchical grouping to optimize an objective function. *Am. Stat. Assoc. Jl* **58,** 236–244.
8. YOUNG FW and TORGERSON WS (1967) TORSCA, A Fortran IV program for Shepard–Kruskal multidimensional scaling analysis. *Behav. Sci.* **12,** 498.

ADDRESS FOR CORRESPONDENCE: *Professor JG San Miguel, Graduate School of Business Administration, Harvard University, Soldiers Field, Boston, Massachusetts* 02163, *USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Organization Structure and Cooperative Market Relations

HELMY H BALIGH

RICHARD M BURTON

(Received September 1975; in revised form April 1976)

A linear model is used as the vehicle for analyzing some of the interactions between an organization's cooperative market relations (product forms) and its structure (decomposition and coordination). Normative implications for organizational choice, and a set of conclusions concerning the nature of organizational change and growth for practical use are developed. The concept of product form as a description of cooperative market relations is the basis for relating the activities of the firm to its market environment. The decision on product form is viewed as a decision on the resource usage vector, and hence as a decision on the nature of the linear programming model formulated. The decision on the decomposition of this linear model and on the solution procedure are viewed as decisions on organization structure. They are viewed as decisions on the partitioning of the organization's decision variables into divisions and on the mechanisms used to coordinate these divisions.

Vol. 4
1976

THE DECISIONS ON DECOMPOSITION, COORDINATION, AND PRODUCT FORM

DECISION problems on certain aspects of organization structure have been widely and conveniently discussed in the context of the decomposition and solution procedures of a linear programming model [1, 3, 5, 8]. The procedure followed in solving a linear model determines elements of the organization's structure viz. the way the organization is decomposed, and the mechanisms used for its coordination [1, 3, 5, 8, 10]. These analyses, however, have started with a given linear model and concentrated only on the solution procedure.

In this paper we consider both the decision on the linear programming solution procedure and the decisions on the nature of the model and modifications in it. Whereas the decision on a solution procedure is a decision on organization structure, the decision on the nature of the model formulated is a decision on the firm's potential cooperative relations. The latter is a decision on

The authors are grateful for the help they received in making revisions of the paper from several of their colleagues in GSBA.

relations of the organization to its environment. Our analysis of these relations is thus a limited focus variation of the general analysis of the relation between structure of the organization and its environment. The literature on this topic includes the work of Thompson [11, 12] which focuses on the relation of uncertainty to decision structures. Lawrence and Lorsch's [9] contingency models focus on relations between environmental uncertainty and internal structure and degree of decentralization. Burns and Stalker's [4] investigation of electronics firms demonstrates the influence of technological uncertainty and rapid innovation on structure. Finally there is Chandler's [6] historic analysis of the general relations between structure and strategy.

Decisions on the organization's cooperative market relations, i.e. on the nature of this one boundary between it and its environment, we term decisions on product form [2]. Product form is discussed in terms of the resource amounts necessary to obtain it. It is a concept that is easily and clearly developed in terms of the linear model, and represents a description of the firm's cooperative relations with its environment. Decisions on product form are decisions on the nature of the linear programming model, and decisions on organization structure are decisions on the decomposition and solution procedures of this model. It is the relations of these two kinds of decisions that is the topic of the paper.

The practical significance of the paper lies in its contribution to comprehensive planning for an organization. It points out some of the needs of organizational structure which arise from changes in the firm's relations with its market environment. It identifies the effects which the existing organizational structure has on the development of changes in the firm's relations with its markets. By showing the nature of the interconnection between the two decisions on organization structure and product form the paper identifies general ramifications of each alone. It supplies relations that must be specifically identified if planning is to become more comprehensive. These practical implications will become more apparent through our use of a simple example. We shall use the case of an organization which manufactures and sells two product lines—tires and batteries, to illustrate our analysis of the possible effects of changes in the product form in each line on organization structure and vice versa.

The analysis uses the linear model of the firm which is:

$$\max \quad \sum_{j=1}^{J} S'_j x_j$$

$$\text{s.t.} \quad \sum_{j=1}^{J} a_{kj} x_j \leq A_k \qquad k = 1, \ldots, K.$$

$$x_j \geq 0 \qquad j = 1, \ldots, J.$$

where $$S'_j \equiv S_j - P_j - \sum_{i=1}^{I} C_i b_{ij},$$

and

x_j ≡ number of units of product j;
a_{kj} ≡ amount of fixed resource k required for a unit of j;
b_{ij} ≡ amount of variable resource i required for a unit of j;
A_k ≡ upper bound of fixed resource, k, $k=1,\ldots,K$;
C_i ≡ unit variable cost for input, i;
P_j ≡ entry price, or purchase price, for process, j;
S_j ≡ exit price, or selling price, for process j;
S'_j ≡ net contribution to profit, $j=1,\ldots,J$.

Further, let $\mathbf{a}_j = (a_{1j}, a_{2j}, \ldots, a_{kj}, \ldots, a_{Kj})$ and $\mathbf{b}_j = (b_{1j}, b_{2j}, \ldots, b_{ij}, \ldots, b_{Ij})$.

The vector $(S_j, P_j, \mathbf{a}_j, \mathbf{b}_j)$ is associated with the jth process or activity. The jth product's form is described by such a vector, and the firm's cooperative market relations may thus be described by a set of vectors. Any change in these cooperative market relations or product forms would be reflected by changes in the set or in the components of its vector elements.

An initial formulation of a linear model entails a decision on a given set of potential product forms which the firm is to consider. The solution of the model involves picking a subset of these to implement. The decision on the process by which this linear model is solved repeatedly to meet parameter changes involves decisions on the decomposition of the model and on the mechanisms to use to coordinate the segment solutions. These decisions are decisions on organization structure. Alterations in potential cooperative market relations mean changes in the form of the linear model to be solved and hence possible changes in structure. Table 1 contains a summary of the relations between types of product form sets and organization structures.

Vol. 4
1976

SEPARABLE PRODUCT FORMS

In the case of independence or separable product forms, the set of product forms and the related linear model are given by the figure obtained from Fig. 1

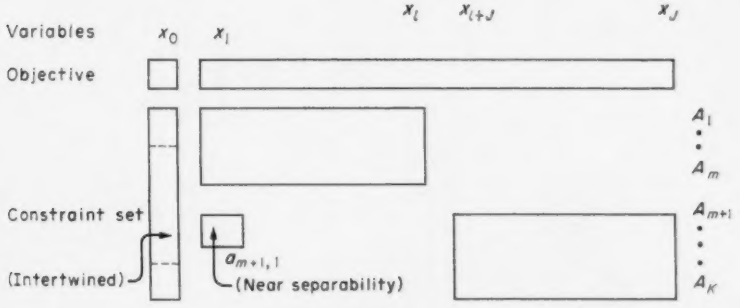


FIG. 1.

TABLE 1. FEASIBILITY AND OUTCOMES OF ORGANIZATION STRUCTURE GIVEN INTERDEPENDENCE OF PRODUCT FORMS

			Structure			
			Centralized	Natural decomposition	Arbitrary resource allocation for decomposition	Budgeting approach	Pricing approach, i.e. transfer pricing
Technological interdependence of product forms associated with a chosen set of cooperative market relations	Separable	a	feasible	feasible	feasible but unnecessary	feasible but unnecessary	feasibly but unnecessary
		b	optimal	optimal			
		c	costly solution procedure	less costly solution procedure			
		d	inefficient	efficient			
	nearly separable	a	feasible	not feasible	feasible	feasible	feasible
		b	optimal solution		nonoptimal solution	optimal solution	optimal solution
		c	costly solution procedure		less costly solution procedure	costly solution procedure	costly solution procedure
		d	efficient (?)		efficient (?)	efficient (?)	efficient (?)
	Intertwined: or non-separable	a	feasible	not feasible	feasible	feasible	feasible
		b	optimal solution		nonoptimal solution	optimal solution	optimal solution
		c	costly solution procedure		less costly solution procedure	costly solution procedure	costly solution procedure
		d	efficient (?)		efficient (?)	efficient (?)	efficient (?)

Entries refer to the following: a. Feasibility of structure; b. Nature of model solution produced by structure; c. Cost of solution procedure; d. Efficiency of structure.

when we ignore altogether the boxes marked (near separability) and (intertwined). This is a case, for example, of a firm that sells tires $\{x_1,...,x_l\}$ and batteries $\{x_{l+1},...,x_J\}$.

The J variables can be partitioned such that variables 1 through l use inputs 1 through m only; and similarly, variables $l+1$ through J use inputs $m+1$ through K only. The set of potential cooperative market relations chosen is that reflected by the particular processes, $x_1,...,x_J$, and the forms of their corresponding vectors $(S_j,P_j,\mathbf{a}_j,\mathbf{b}_j)$.

This formulation of the problem is naturally decomposable into 2 problems, tire problems and battery problems. Let division 1 be assigned responsibility for $\{x_1,...,x_l\}$ given resources $\{A_1,...,A_m\}$. Similarly, division 2 sets $\{x_{l+1},...,x_J\}$ given resources $\{A_{m+1},...,A_K\}$. The coordination issue between divisions 1 and 2, tires and batteries, is non-existent—they set no dependent variables and share no common resources.

Consider a change in product form of a simple kind, e.g. it is proposed that $S_j \rightarrow S_j + \Delta S_j$ and that $a_{ij} \rightarrow a_{ij} + \Delta a_{ij}$. More cord, or more rubber could be added to a tire. Or, thicker plates are put in a battery. If both ΔS_j and $\Delta a_{ij} > 0$, then the change involves putting a larger effort into the product and receiving a larger selling price for it. This product form change could be thought of as a 'quality' change. We suggest that minor adjustments in quality, not involving new technology, have minor, if any, organizational implications. For any proposed change, two issues arise: first, is the change desired, and second, does the change in product form require an organizational change as well?

The proposal that $a_{ij} \rightarrow a_{ij} + \Delta a_{ij}$ brings about a trivial change in formulation, for $i \leq m$, and the question on the organization structure has an obvious answer. There is no change in the organization structure. The firm expands the model to include the alternative process given by: $(S_j + \Delta S_j, P_j, (a_{1j},...,a_{ij} + \Delta a_{ij},...,a_{mj}, 0,...,0,...,0), \mathbf{b}_j)$, and solves the new but still naturally decomposable problem. A marginal approach, not involving a new total solution would be to compare the marginal revenue, ΔS_j, with the marginal cost, $\Delta a_{ij} \cdot \lambda_i$ where λ_i is the shadow price (dual variable value) of the current optimal solution. Only if the marginal revenue exceeds the marginal cost, is the proposed product form change desirable.

NEAR SEPARABILITY

The nearly separable case is more complex. It is described by the figure obtained from Fig. 1 when only the box marked intertwined is ignored. Consider the 1st process for division 1, and let $S_1 \rightarrow S_1 + \Delta S_1$ be the price change and $a_{m+1,\,1} = 0 \rightarrow a_{m+1,1} > 0$. This change eliminates the previous simplicity, and the natural decomposition; the first division changes its product such that it

Vol. 4
1976

requires the use of a resource, $m+1^{st}$, which is currently used by, and the responsibility of, the second division.

Ignoring the organizational issue for the moment, we could proceed as before to analyze the profitability of the proposed change by either incorporating the proposal into the total firm model, or by comparing the marginal revenue ΔS_1 for the new tire with the marginal cost, $a_{m+1,1} \cdot \lambda_{m+1}$. The first calculation is not possible with a naturally decomposed structure, although the second calculation is. If the change is desirable at this level of analysis, i.e. $\Delta S_1 > a_{m+1,1} \cdot \lambda_{m+1}$ the next step is implementation and that involves changing the organization structure and incurring further costs.

It is possible to ignore the fact that the $m+1^{st}$ resource used by divisions 1 and 2 is really the same, and thus, the tire division would procure its own electrolysis setup, and the battery division would remain with A_{m+1}. This is an arbitrary resource allocation structure. The change in cooperative market relations is associated with a particular change in formulation of the model. Clearly, this reduces the coordination cost, but would increase the resource cost as it would only be by accident that an optimal allocation of the $m+1^{st}$ resource, electrolysis capacity, would be independently attained. Galbraith's [7] "creation of self-contained tasks" is an identical approach.

Other organization structures to be considered may be resource guided (budgeting) or price-guided (transfer pricing) or centralized. In either of the first two cases, it is necessary to set up a central coordinator who must consider the best allocation of the $m+1^{st}$ resource between the divisions. The centralized organization has been considered previously, and there remains a comparison among only the other two.

The resource guided organization [5, 8] would operate in the following manner. An initial allocation of the $m+1^{st}$ resource would be made to divisions 1 and 2. Using this trial allocation, they would then make plans and determine the marginal value of the $m+1^{st}$ resource, and report the value back to the central coordinator. The central coordinator would adjust the allocation giving the division with the higher marginal benefit a little more, the other a little less. Continuing this procedure, it is sufficient for an optimal allocation that the central coordinator terminate the process when both divisions obtain the same marginal benefit from the resource. Clearly, this planning process is costly. There is the required central coordinator. Each division must plan at least once, and possibly for a large number of trials. The number depends critically upon how 'good' the initial trial allocation is. The closer it is to the best allocation, the fewer trials required. This process is costly and time consuming.

The price guided system [3] would proceed as follows: The central coordinator 'owns' the $m+1^{st}$ resource and sells it to the divisions. The central coordinator would compare the demand against the amount available. If the demand exceeds the availability, he would raise the price; and vice versa. The central coordinator would stop where the demand did not exceed the avail-

ability (i.e. the market is cleared). This system could involve a large number of iterations, depending upon how well the price is set. Our remarks concerning the trade-offs in the resource guided system apply here as well.

A question remains concerning the relative desirability of the resource guided system *vis a vis* the price guided system. This is an empirical question concerning their relative efficiencies and the answer depends upon the particular situation.

INTERTWINED PRODUCT FORMS

The simplest intertwined situation would involve a product say x_0, whose production requires inputs from all K resources, i.e. $(S_0, P_0, \mathbf{a}_0, \mathbf{b}_0)$ is such that $a_{k0} > 0, \forall\, k$. The problem is described by the whole of Fig. 1. For the tire and battery firm, a product using all the resources of both divisions may be a battery covered in a tire rubber casing for protection.

Here, the array of possible organization structures is the same as the nearly separable situation and the analysis of the previous section applies. However, the relative efficiences of different structures may not be the same. It is quite clear that in this case the central coordinator has a larger problem—he has either more resources to budget, or more transfer prices to set. In either of the budgeting or transfer price situations, it is likely that this set of cooperative market relations will involve more costly coordinated solution procedures, and higher returns to such procedures than will the set of nearly separable cooperative market relations. As a result, the problem of finding the best structure in the intertwined case can be expected to be more critical to the organization's well being than in the nearly separable case.

CAPITAL BUDGETING, COOPERATIVE MARKET RELATIONS AND ORGANIZATION STRUCTURE

In the separable product form section, we considered the change in product form reflected by $\Delta a_{ij} \neq 0$ with a price change ΔS_j. As a short-run consideration, we require $\Delta S_j > \Delta a_{ij} \cdot \lambda_i$ for the change to be profitable. Thus, for a single period of unit length the differential cash flow is $\Delta S_j - \Delta a_{ij} \cdot \lambda_i$. Let us assume this obtains for all time periods. Then, the present value of this change is:

$$\sum_{t=0}^{T} \frac{\Delta S_j - \Delta a_{ij} \cdot \lambda_i}{(1+r)^t}.$$

Vol. 4
1976

Where T is the life of the new organization and r is the discount rate. This present value must be higher than any one time cost required in changing the product form or the organization's structure.

Let R be the one time cost of changing the organization structure and product form of x_1, giving us the near separability form. The differential cash flow is

$$\Delta S_1 - a_{m+1,1} \cdot \lambda_{m+1} - w$$

where w is the recurring added operating costs for the new organization. Then, if the present value

$$\sum_{t=0}^{T} \frac{\Delta S_1 - a_{m+1,1} \cdot \lambda_{m+1} - w}{(1+r)^t}$$

is greater than the initial cost R, the change is profitable. Clearly, there exists a different set of one time costs R and differential operating costs w for each organizational structure, whether it is centralized, arbitrarily decomposed or decomposed and coordinated by a budgeting or pricing approach. These form a set of exclusive and exhaustive structures from which to choose one to accompany a product form change. The best choice is the one for which the present value exceeds the initial costs by the greatest amount. Thus, for each product form the best structure over the long run is identified, and the information necessary to compare different pairs of product form change and best organization structure is supplied. It is our observation that capital budgeting analysis of new products frequently ignores organizational issues and the costs involved in making organizational changes, if they are required.

ORGANIZATION STRUCTURE AND THE POTENTIAL SET OF COOPERATIVE MARKET RELATIONS

Knowledge of the complexity of solving certain kinds of formulated linear models may be used in the process of model formulation. The anticipated needs of structures may be used to alter the set of potential cooperative market relations in the initial formulation.

In the process of choosing a structure, one may deliberately remove potential cooperative relations from consideration, i.e. change the initial formulation, since the presence of some such elements in the set may make solution procedures of all types (all structures) much more *costly*. The decision to remove such cooperative market relations may well precede any attempt at model solution. For example, our tire and battery firm produces two sets of products that allow for a natural decomposition of the production divisions, i.e. car tires and batteries of various kinds and sizes. The retail marketing of the two sets of products in the organization's own retail stores involves the use of resources common to both, as suggested in Fig. 2, e.g. resources $A_{r+1},\ldots,A_k$. These resources, involving distribution, retail advertising, space, etc., are many in

Vol.
197

number, and the coordination of their use by both products involves high costs. When the two sets of products are sold only at wholesale, these resources are few or zero in number and the costs of coordinating the marketing efforts of the two products are low. Any advantages to retailing of the products may well be more than counter-balanced by the increased costs of coordination. The firm may choose to stay out of the cooperative market relations involved in retailing. By so choosing it has allowed the requirements of structure to determine its cooperative market relations.

The organization structure and implemented set of cooperative market relations have an effect on the area which elements of the organization *search* for alterations of potential cooperative market relations. Suppose the firm had found it profitable to sell car tires and batteries at retail in its own stores, and had created a division to coordinate the marketing of both kinds of products. Under these conditions the retail stores division would seek new cooperative market relations in the form of new products which would improve the assortment of goods in its retail stores. It might seek automobile accessories to sell in its retail stores. If the firm had chosen to sell its car batteries and tires at wholesale, the car tire division may seek to use its expertise in its new cooperative relations and maintain its independence of the battery division by manufacturing airplane tires. In this case the absence of the retailing division reduces the pressure in the firm for the search for products compatible with *retail marketing* of tires and car batteries. The search and pressures are much more likely to be for those products compatible with the *production* of car batteries and car tires.

Vol. 4
1976

Independent divisions may be expected to seek changes in product forms that maintain their independence. Further, the absence of a central coordination unit is likely to mean that no pressure exists to expand the interdependence between divisions, and hence the operations of such a central unit.

Suppose the formulation were the intertwined one described by Fig. 1. Here

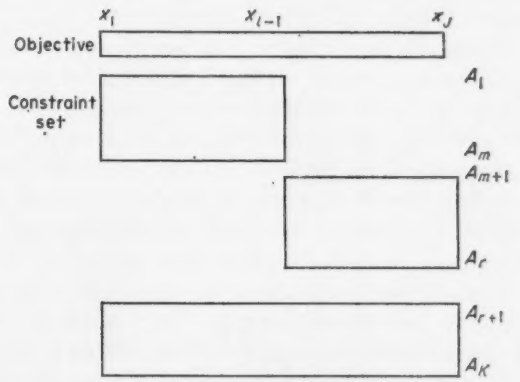


FIG. 2.

a centralized structure is costly, and a decentralized structure with x_0 in either of two divisions or comprising a division by itself is likely to involve many solution iterations for coordination. A decentralized structure without coordination, i.e. a solution procedure that produces a feasible but non-maximizing solution, may be best. Whatever estimate one may have regarding the best structure for this formulation one should consider the possibility of altering the formulation by removing x_0 from the set of potential cooperative market solutions altogether. This would yield a set of separable product forms. Thus, using the procedures outlined earlier a decision may be made to formulate the simpler model, leaving out of consideration the cooperative relations involved in x_0 and to choose a decentralized structure given that certain parameters, e.g. price of x_0, are within certain ranges. It may also be planned to alter the formulation and the structure if the product x_0's revenues and variable costs enter other ranges.

One can conclude that, because of its effects on structure, a product form such as that given by x_0 in Fig. 1 is not likely to be organizationally viable unless it is a *major* contributor to the firm's profit. The larger the number of such products the more likely that any one will be viable given any contribution it makes to profits.

If the initial formulation had a set of product forms with a large set of shared resources (markets) of the form described in Fig. 2, and the structure were centralized, then the central decision unit is likely to search for forms that would improve the use of the shared resources even if it added complexities by removing a partial natural decomposition. The addition of x_0 product forms is more likely to follow from the structure given in Fig. 2. Such forms are those to which the central decision unit owes its existence, and which involve it in a 'least learning' expansion process. The extent to which the new form will remove partial natural decompositions may well be expected to depend on the political power relationships between the central coordination unit and the divisions representing the partial natural decomposition.

Not only is the area of search likely to be dependent on structure as a result of the behavior of individuals, but formulations turned up by the search will be compared with an eye to the fact that structure changes tend to involve large setup costs. To go from a centralized structure to a decentralized one requires people to be involved in computation, whereas before they had not been so engaged. To go from less coordinated decentralized to more coordinated decentralized structures requires an increase in the information processing done by individuals, and an increase in their number, and in the specificity of the decision rules to which they must adhere. Structural inertia may specify future market developments unless these are explicitly planned. Meanwhile planned future market developments cannot ignore these structural requirements, and the costs they may entail. These are the costs of re-educating decision makers and overcoming structural inertia.

Vol.
197

SUMMARY

The paper develops a conceptually operational view of both structure and the cooperative market relations that makes the two decision problems compatible for simultaneous treatment. The vehicle used is a tire and battery firm viewed as a linear activity analysis model. The decision on cooperative market relations is cast as one involving a choice of a potential set of product forms. The choice is equivalent to the choice of the formulation of a linear model of the firm, or some small set of such formulations. From this set of potential relations there will emerge a subset of cooperative relations which the organization will choose to have with its market environment.

The particular manner in which a given model formulation affects the costs and returns to various structures has been analyzed. These costs and returns are those of different methods of solution, i.e. structures. The manner in which a structure may be used to choose from a set of potential cooperative market relations (model formulations) has been identified. Using a capital budgeting framework, we are able to suggest a normative criterion for the comparison of proposed changes in product form and organization structure and in long run capital expenditure on resources fixed in the short run. Finally the demands of structure costs, i.e. solution procedure costs, on problem formulations and hence on the potential and therefore chosen cooperative relation sets has been investigated and identified.

Vol. 4
1976

REFERENCES

1. Atkins D (1974) Managerial decentralization and decomposition in mathematical programming. *Opl. Res. Q.* **25**(4), 615–624.
2. Baligh HH and Richartz LE (1967) *Vertical Market Structures.* Allyn and Bacon, Boston.
3. Baumol W and Fabian T (1964) Decomposition, pricing for decentralization and external economies. *Mgmt Sci.* **11**(1), 1–32.
4. Burns T and Stalker GM (1961) The management of innovation. Tavistock, London.
5. Burton RM, Damon WW and Loughridge DW (1974) The economics of decomposition: resource allocation vs. transfer pricing. *Decis. Sci.* **5**(3), 297–310.
6. Chandler AD, Jr (1962) *Strategy and Structure: Chapters in the History of the American Industrial Enterprise.* MIT, Cambridge, MA.
7. Galbraith Jay (1973) *Designing Complex Organizations.* Addison Wesley, Reading, MA.
8. Kornai J and Liptak Th (1965) Two-level planning. *Econometrica* **33**(1), 141–169.
9. Lawrence PR and Lorsch Jay W (1969) *Developing Organizations: Diagnosis and Action.* Addison-Wesley, Reading, MA.
10. Marschak J (1955) Elements for a theory of teams. *Mgmt Sci.* **1**(2), 127–137.
11. Thompson JD (1964) Decision-making, the firm, and the market, in *New Perspectives in Organization Research* (Eds. Cooper WW, Leavitt HJ and Shelly MW). John Wiley, New York.
12. Thompson JD (1967) *Organizations in Action.* McGraw-Hill, New York.

ADDRESS FOR CORRESPONDENCE: *Professor HH Baligh, Graduate School of Business Administration, Duke University, Durham, NC* 27706, *USA.*

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Design of a Strategic Planning Management Information System

KYUNG II GHYMN

WILLIAM R KING

(*Received April 1976; in revised form May 1976*)

Information systems that are to support strategic planning must be developed on bases that reflect the unique nature of strategic organizational decisions. This paper presents a systems design methodology which implements five basic criteria for the design of a strategic planning Management Information System (MIS). The methodology uses discriminant analyses of the perceived importance to managers of various categories of strategic information to delineate the most significant information requirements of various managerial groups. These results are augmented with similar analyses based on the personal characteristics of the potential system users in order to implement a design criterion related to the integrated nature of a planning MIS and a 'teaching' objective which is posited to be valid and feasible for such a system.

Vol. 4
1976

A METHODOLOGY FOR THE DESIGN OF A STRATEGIC PLANNING MIS

THE PROBLEMS that are inherent in developing an information system which is to serve managers in support of their strategic planning responsibilities have been widely discussed. Some, such as Wrapp [18], argue that since strategic problems are unique, there is little possibility for specific systematic information support of strategic planning. Others, such as King and Cleland [8] and Kotler [11] have provided general designs for strategic planning information systems and referred to the effectiveness achieved by those business firms which have developed them.

Despite the controversy and the success of some strategic planning management information systems (MIS), there exists no generally-accepted methodology for translating strategic organizational attributes such as the organization's objectives and top management's perceptions of their information needs into

the design parameters for a strategic planning MIS. Although this methodological void pervades the entire field of MIS design, it is particularly important at the level of strategic planning because this level of management directly involves important unstructured decision processes which could benefit greatly if timely and relevant information were available to strategic managers.

This is so because strategic decisions inherently involve more external considerations than do tactical or operational decisions, and because the information on which they must be based is heavily biased toward environmentally-generated, as opposed to internally-generated, data. This essential environmental orientation for a strategic planning MIS complicates the design process because the strategic planning MIS must necessarily involve diverse data, multiple data sources and wide variations in feasible system configurations.[1]

Many of the existing approaches to the development of management information systems were developed in operational-level decision-support contexts. Even those which might be thought of as being 'strategic' in their orientation, such as that of Ackoff [1] and King and Cleland [9], do not adhere to some of the design criteria which would appear to be critical to the development of an effective strategic planning MIS.[2]

This paper presents a methodology for the development of the crucial 'information requirements' dimension of a general MIS design. The specification of this, "Who needs which information?", element of an MIS design is one of the most troublesome elements in designing systems to support all levels of managerial decision making.[3]

This strategic planning MIS general design[4] methodology has been developed and evaluated on an experimental 'pilot test' basis in a large US-based, multinational corporation with emphasis on international strategic decisions and internationally-oriented managers. The international domain is most appropriate as a milieu for demonstrating such an approach because there is a clearer line of demarcation between 'internal' and 'environmental' data and between 'new information' and basic 'managerial knowledge' in this context than in most others. Also, in the international domain, managers are believed to be much more likely to recognize that they need environmental information[5] and to be more conscious of their efforts to obtain it.[6]

[1] See King and Cleland [8] for a detailed discussion of environmentally-oriented information systems.

[2] These criteria are discussed in the next section.

[3] See King and Epstein [10] for a review of the various approaches which have been taken to the determination of information requirements in MIS design.

[4] The term 'general design' is used synonymously with 'conceptual design,' and 'gross design' to mean the establishment of general performance requirements and specifications for the MIS. See Murdick [12].

[5] See Cyert, Dill and March [6].

[6] See Taylor [16].

CRITERIA FOR THE GENERAL DESIGN OF A STRATEGIC PLANNING MIS

If strategic decisions requiring extensive environmental information are to be supported by formal information systems, there is a clear need for the objective determination of information requirements. Strategic decisions require timely information and many must be made in a time span that will not permit the extensive collection of basic data on an *ad hoc* 'study' basis. Thus, for a 'system' to be useful, it must possess a data base which can be drawn on for relevant strategic information.

To be effective and efficient, the strategic planning management information system which is to be developed around such a data base must meet a number of other general design criteria. First, the data base must reflect, and the system must provide, information which satisfies the perceived needs of its user–managers. A variety of studies have reached this general conclusion[7] despite the fact that the difference between manager's 'wants and needs' has been widely discussed in the literature of information systems.[8]

Vol. 4
1976

The second criterion relates to cost–effectiveness in that even a user-oriented system cannot be designed to idiosyncratically provide the specific information desired by each manager. To provide such a capability would require excessive development cost and continuous system design changes as managers changed positions and people entered and left the organization. Thus, if a user-oriented system is to be cost-effective and stable, it must provide various classes of managers with specified categories of information.

The third criterion is parsimony. This criterion implies that the system will be designed to provide specified managers with reports containing that critical information which is most relevant to the support of their strategic decisions. Of course, an inquiry capability may be incorporated into the system to allow for the broader needs of some individuals under certain circumstances, but the basic structure of the system must have the objective of providing the most useful relevant information without inundating the manager with masses of less useful data.

The fourth criterion relates to the inherent 'integrated' nature of strategic decisions and of the information which is most effective in supporting such decisions. This strategic information necessarily involves many areas of the organization's activities as well as many diverse organizational clientele groups and areas of environmental influence. Thus, a strategic MIS will generally provide managers with information which goes beyond the domain of their specific operating responsibilities. This design criterion recognizes the distinction

[7] See Chandler [3], Cyert and March [6], Weick [16], Child [4], and Normann [13].
[8] See Ackoff [1] and King and Cleland [9].

between strategic and operating decisions and the fact that the information required for one variety of decisions may be inadequate for the other.

The fifth criterion, superficially somewhat at divergence with the previously stated criteria, has to do with the validity of a 'teaching' objective for a planning MIS. Particularly in areas such as strategic planning and the international business arena, there is potentially-great value in providing managers with more, or different, information than they might personally request for any given situation. This is so because of the high turnover in management personnel in these areas and because managers are often thrust into them without the benefit of formal training [2].

Thus, although managers' perceived needs do provide a valid basis for system design, those perceived needs must reflect the collective wisdom of some relatively homogeneous group of experienced managers rather than the needs as perceived by any individual or by any group of inexperienced managers. This concept of information-needs perceptions based on the collective wisdom of experienced managers is consistent with the other criteria. In addition, it creates the opportunity for the development of an information system which provides a manager with a set of information which has been assessed to be 'needed' by a 'typical' experienced person occupying each organizational position, rather than for the specific individual who occupies each position. This cost-effective approach has the added benefit of enabling 'the system' to educate the novice manager by providing information which he might not personally be able to identify as important, or even relevant [10].

Vol.
197

DESCRIPTION OF THE MIS DESIGN METHODOLOGY

The proposed strategic planning MIS design methodology will be described in the context of the real-world situation in which it was developed and tested. The basic objective of the methodology is the objective delineation of relatively homogeneous classes of managers and the specification of the categories of information which are deemed to be most relevant to them. If relative homogeneity exists for various managerial categories with regard to the perceived relevance and importance of information, such a structure forms the basis for a general MIS design.

The company selected for the demonstration is a large US-based manufacturing conglomerate which operates in 34 countries. A set of 54 company managers who were identified as being 'US-based international managers' through the analysis of organization charts and job descriptions[9] and through discussions with senior managers agreed to participate in the study.

[9] US-based managers were chosen because they are hypothesized to be more homogeneous, because they would be the initial users of any MIS to be developed, and for the other reasons discussed earlier.

The design methodology involves the administration of a survey instrument to prospective user–managers. In this instance, the instrument contained queries related both to personal and organizational attributes of the respondents and to their perception of the degree of importance of a wide variety of environmental information.

The personal and organizational attribute data which were collected from each manager were used to define various 'managerial categories' for the study. Eight attributes of the individual/position were used (parenthetical descriptions describe the levels of each attribute as used in this situation):

1. Hierarchical level (top, middle and lower levels).[10]
2. Functional area (marketing, finance, production, engineering and planning).
3. Product line group (energy products, industrial products, consumer products and services).
4. Line *vs.* staff.
5. Overseas residence experience (overseas residence assignment experience *vs.* no experience).
6. Number of overseas business trips (no trips, 1–10, 11–20, 21 or more).
7. Length of service (2–10 years, 11–20, more).
8. Education background (business–economic *vs.* science–engineering background).[11]

Vol. 4
1976

The first four attributes have to do with organizational position while the latter four are personal in nature. The personal characteristics would undoubtedly not be used as bases for most general MIS designs. For instance, managers would not normally be provided with different information as a function of their length of service. However, the potential 'teaching' objective for a strategic planning MIS, as well as the criterion related to experienced managers serving as a basis for information requirements determination, may well be fulfilled by analyses of differences in information perceptions in terms of the latter four personal characteristics.

The questionnaire distributed to the managers presented environmental information delineated on the basis of 38 environmental variable categories

[10] Top level includes: Directors of divisions, executive staffs, General Managers, Vice Presidents, Executive Vice Presidents and above.
Middle level includes: Department managers, district managers, project managers, senior consultants (internal), senior staffs.
Lower level includes: Area managers, assistant managers, section chiefs, consultants (internal) junior level staffs.

[11] A third level of this characteristic—'humanities'—was originally proposed but no respondents fell into this category.

which were identified by Piper [14] as significant to foreign investment and business operations decisionmaking in studies of several large multinational companies and 16 investment situations. These variables are shown in Table 1 where they are grouped into six broad categories—'economic/legal', 'financial', 'social', 'political', 'marketing', and 'technical/engineering'. The questionnaire asked the managers to rate the importance of each of the informational variables on a Likert-type five point scale ranging from 'very important' to 'rarely important'.

TABLE 1. ENVIRONMENTAL VARIABLES IN FOREIGN INVESTMENT AND BUSINESS OPERATIONS DECISION MAKING*

Economic and legal
1. Legal systems of host country
2. Host government attitude toward foreign investment
3. Demand and supply conditions for the product
4. Restrictions on ownership
5. Tax laws
6. Import/export restrictions
7. GNP/per capita income
8. Inflation
9. Level of industrialization

Financial
10. Capital availability
11. Acquisition and merger
12. Projection of cash flows
13. Return on investment
14. Monetary exchange
15. Insurance against risks (expropriation, nationalization, etc.)

Social
16. Social unrest
17. Religion/language/racial barriers
18. Labor organizations
19. Public literacy
20. Public attitude toward foreign investment
21. Living conditions for American managers and their families (schools, personal security, cultural difficulties, etc.)

Political
22. Host government political system
23. Political instability
24. Relations with neighboring countries
25. Relations with supra-national organizations (UN, EEC, IMF, GATT, OECD, World Bank, etc.)
26. Political party factions
27. Attitude of political opposition toward foreign investment
28. Military elite power in politics
29. Communist influence

Marketing
30. Market potential
31. Competition
32. Distribution channel systems
33. Production cost and pricing
34. Social/cultural factors impacting upon products

Technical and engineering
35. Technology and its transferability
36. Raw material availability
37. Availability of cheap labor and trained management
38. Infrastructure to support business (electricity, telephone, roads, ports, etc.)

* Adapted from Piper [14].

The proposed MIS design methodology involves the statistical analysis of these responses to determine groups of managers who are relatively homogeneous in their assessments of informational relevance and to identify categories of variables in terms of their relative importance. The basic statistical technique used in the methodology is multivariate discriminant analysis. Analyses were

carried out by the Clyde–Cramer approach using the Wilk's Lambda criterion [5, 15]. This program of multivariate discriminant analysis scans all variables and then selects those which contribute most to the separation of the predefined managerial groups. The important points in the use of this tool in this methodology for the determination of 'information requirements' are: first, the discriminant functions are chosen in such a way as to minimize within—group variance relative to between—group variance; second, to determine the order of importance of the variables, discriminant functions are normalized so that their variance in each of the groups is unity; third, only statistically significant discriminant functions are used; fourth, a univariate F test is used in order to provide a basis for examining differences with respect to each of the individual variables by given managerial groups. Finally, the extent of usefulness of a given discriminant function coefficient depends not only upon the reasonableness of the variables selected and the percentage of discriminable variance for which this function-coefficient accounts, but also on the group mean vectors in the discriminant space (the extent of the 'distance' between the group means of the separated groups). For this purpose, the group mean vectors are analyzed to identify which management groups differently perceive each class of variables.

Vol. 4
1976

Hierarchically-defined managerial group analysis

To demonstrate this analytic procedure, consider managerial groups defined solely by hierarchical level. In applying discriminant analysis to survey responses grouped in this way, one is asking the questions: "To what extent do managers in different hierarchical levels (top, middle and lower level managers) perceive various external environmental variables to be important?", and "Is there a significantly great difference in their perception to permit these perceptions to be used as a basis for the design of an MIS according to the previously-stated criteria?"

Table 2 indicates that in this firm, the largest contributor to hierarchical group separation in the first discriminant function is the Political variable. The Wilk's lambda value for the three groups is 0·899, which is significant at the 0·001 level. An examination of the group mean vectors (Table 3) shows that the mean value on the Political variable by top managers is 4·42, whereas middle and lower level managers have mean scores of 2·71 and 2·01, respectively. Since the likelihood that such a difference would be due to chance is less than one in a thousand, the null hypothesis that the three management classes have similar perceptions with regard to the external variables can be rejected.

Univariate F tests for statistical significance of group differences with respect to each of the six variables provide additional information about managers' perceptual differences. For example, Table 4 shows that managers share similar perceptions with respect to the Marketing, Technical, and Financial variables while they have sharp differences in their perception with respect to the Political, Economic, and Social variables. For the latter case, again referring to Table 3,

TABLE 2. HIERARCHICAL LEVEL DISCRIMINANT ANALYSIS

Variables	Standardized discriminant function coefficients*
Economic and legal	0·348
Financial	0·378
Marketing	−0·129
Political	0·900
Social	0·392
Technical and engineering	0·005

$\Lambda = 0{\cdot}899, P < 0{\cdot}001$.

* The number of independent discriminant functions that can be derived when the number of variables exceeds the number of groups is one less than the number of groups. For example, there are six variables and three groups (top, middle and lower level managers) for this case. Therefore, only two independent discriminant functions can be derived. However, only the first discriminant function is given here because this is the only function that turned out to be statistically significant. In the other analyses reported here, only functions which are statistically significant at the (0·05) level are used.

TABLE 3. HIERARCHICAL LEVEL MEANS AND (STANDARD DEVIATIONS) ($N = 54$)

Variables	Top ($N = 16$)	Middle ($N = 21$)	Lower ($N = 17$)
Economic and legal	3·91 (0·73)	3·29 (0·86)	2·50 (0·81)
Financial	3·94 (0·75)	3·74 (0·84)	3·22 (1·27)
Marketing	3·24 (0·98)	3·26 (1·23)	3·11 (1·37)
Political	4·42 (0·35)	2·71 (0·79)	2·01 (0·41)
Social	3·16 (1·00)	3·11 (0·92)	2·07 (0·94)
Technical and engineering	3·13 (0·99)	2·90 (1·06)	2·98 (1·37)

TABLE 4. UNIVARIATE *F*-TESTS, HIERARCHICAL LEVEL

Variable	$F(2, 51)$	Mean SQ	*P* Less than
Economic and legal	12·654	8·288	0·001
Financial	2·419	2·293	0·099
Marketing	0·083	0·120	0·921
Political	76·287	25·261	0·001
Social	7·266	6·616	0·002
Technical and engineering	0·175	0·231	0·840

it is found that top and middle level managers regard Economic and Social variables as more important than do lower level managers, and with respect to the Political variable, it is the top-level managers who regard it as the most important, whereas the middle and lower level managers regard it as relatively less important.

Thus, on the basis of this hierarchical managerial group analysis, the MIS designer may tentatively conclude that the political information is regarded as most important by top-level managers and that there are sharp differences between upper and lower level managers with regard to their assessment of the importance of economic and social variables. This provides preliminary guidance to the MIS designer in terms of the critical, "Who gets what information?", question of MIS design.

Functionally-defined managerial group analysis

Using the same methodology, the designer may then turn his attention to the functional dimension of the managerial taxonomy. Since the emphasis here is on the methodology rather than the specific numerical results of the pilot test, we shall not provide more than the illustrative numerical results of Tables 2–4.[12] When similar discriminant analyses were performed using managerial groups defined according to function, it was found that Technical, Engineering and Marketing variables were those that contribute most to the separation of groups on the first discriminant function, while Financial and Social were the contributors on the second discriminant function, and Political was most significant on the third discriminant function. (All three functions were statistically significant.)

The conclusions reached through this functionally-defined managerial group analysis were that engineering and production managers pay greatest attention to Technical and Engineering variables. Marketing and finance managers also present a striking profile of attention to Marketing and Financial areas, respectively. This was reflected in the univariate *F*-tests which revealed that there is a significant difference in perception between functional groups with respect to all environmental variables except the Economic and Legal variable.

This finding supports the hypothesis that a specialist in a functional area, having a limited knowledge outside of his own area, will tend to seek more functionally related information and consider it to be most important. This portion of the pilot test results is certainly not surprising. However, the crucial importance of the finding for MIS design is the objective empirical verification of the intuitive proposition that a person's area of work and responsibility has great impact on the kind of information that he deems to be most important.

If this finding were to be accepted by the systems designer and used, as it often is, as the only basis for the system design, the potential for an integrated strategic planning system would be nullified. However, in the context of the previously-enumerated planning MIS design criteria, these results serve a much more creative purpose in that they suggest the overlapping interests of managers in various functional areas which may be used to help in achieving an 'integrated' system which can perform 'teaching' functions.

[12] Numerical results for this grouping as well as for the others to be discussed are available on request from either of the authors.

Product-group managerial-group analysis

The company in question has three distinctive product groups: energy products, heavy industry products, and consumer products. Each product group has an executive who reports directly to the Corporate Chief Executive. In addition, there are Corporate Planning and Corporate Affairs divisions, both of whose division heads report directly to the Corporate Chief Executive. Since the overall functions of these two divisions are of the variety of management services, they are merged, and named 'Service Group' for the purpose of this analysis. Thus, the 'product groups' are thought of as consisting of an Energy Group, Industry Group, Consumer Group, and Service Group.

The discriminant analysis revealed that Social, Political, and Technical and Engineering are the variables that contribute most to the separation of the product group classification. By examining the mean responses by product group, it was determined that the Energy Group and Industry Group showed high ratings on the Technical and Engineering variable, whereas the Service Group did not. (The Consumer Group was found to be between these two extremes.)

This analysis thus showed that the perceived importance of Technical and Engineering variables serve to distinguish the Energy and Industry Groups from the Service Group, and that the Political variables are paid more attention by the Industry Group—thus indicating other important guidelines for the MIS designer in determining who should be provided with which information.

Line vs. staff analysis

The data may also be similarly analyzed on a line *vs.* staff basis. However, since the Service Group is predominately staff while the other groups are predominately line, this analysis adds little in the case of this pilot study.

Other analyses

Although the other four attributes used in the study—overseas residence assignment, overseas travel, length of service, and educational background—are not to be used as a primary basis for the general design of an MIS, the application of the discriminant methodology to these characteristics may be applied in developing a 'teaching' dimension for the MIS and in applying the design criterion related to the use of a 'typical experienced' manager as a focal point for MIS design.

Analysis on the basis of the four personal characteristics showed that:

(a) Whether or not a manager had previously had an overseas residence assignment principally reflected itself in his perception of the importance of Social variables—those with such experience regarding these variables as quite important.

Vol. 19

(b) Managers who had made large numbers of overseas trips regard Social, Economic, Political and Legal variables as much more important than those without such extensive travel.

(c) Managers with longer service tend to emphasize Political variables as relatively more important than do less experienced managers. (Of course, since there is a high correlation between length of service and management level, this is simply another way of drawing the conclusion regarding top management as discussed earlier.)

(d) Social variables contributed the most to group separation on the basis of educational background, but surprisingly, it was the Science and Engineering majors who regarded these variables as most important. Unexpectedly, no significant differences existed with regard to the educational background groups with respect to Economic, Legal, Marketing, and Technical variables.

These analyses in terms of personal characteristics permit the system designer to augment the results based on analyses in terms of organizational characteristics. For instance, the first three characteristics—overseas residence, overseas trips, and length of service—can serve as operational definitions of 'experience' in applying the systems design criterion that specifies that the system should fulfill the needs of an experienced manager in each position. In this instance, the augmentation of a basic information set with social variables would appear to be indicated.

In a similar vein, if a 'teaching' objective is deemed to be appropriate for the system, the 'educational background' analysis may provide insights. It might be interpreted to suggest a receptivity on the part of technical people to broadening their scope to encompass social variables.

CONCLUSIONS

This paper has presented a methodology for the design of a strategic planning MIS which emphasizes the perceived information needs of managers who are involved in high-level planning activities. A set of criteria for such systems has been proposed and a methodology for implementing the criteria has been shown to be feasible in a pilot study context involving a large multi-national corporation's 'international managers'.

The methodology is focused toward answering the, "Who needs what information?", question which reflects a critical dimension in the development of a general systems design. In the corporation in question, the methodology, as applied to management categories defined in organizational terms, clearly

TABLE 5.

Information category	Management category
Political	Top
Economic, social	Top, middle
Technical and engineering	Engineering, production
Marketing	Marketing
Financial	Financial
Technical and engineering	Energy group industry group
Political	Industry group

demonstrates the importance of the information-management category relationships shown in Table 5.

The methodology also offers the opportunity for complementing these organizationally-based results with the results of analyses based on personal attributes of potential system users.

In this instance, such analyses suggest that experienced managers tend to give greater emphasis to social variables. Since one of the systems-design criteria specified the desirability of defining information needs in terms of those perceived by a 'typical' experienced manager in a position, this suggests that the basic information set for any position might well be augmented with social information.

Similarly, an analysis based on the educational background of the managers suggests a desire for social information by managers with technical backgrounds. This can, therefore, provide a useful basis for implementing a 'teaching' objective for a system, since the horizons of these managers may be broadened in the most behaviorally-acceptable manner through this variety of information.

Taken together, these relationships serve to answer the critical, "Who needs what information?", question of MIS design at a level which is suitable for a general systems design and in a fashion which addressed the crucial issues of information overload, parsimony, and cost–effectiveness which are a part of every systems design process.

REFERENCES

1. ACKOFF RL (1967) Management misinformation systems. *Mgmt Sci.* **14** (4), B147–B156.
2. ANDREWS K (1971) *The Concept of Corporate Strategy*, Chapter 1. Dow Jones Irwin, New York.
3. CHANDLER A (1962) *Strategy and Structure: Chapters in the History of the Industrial Enterprise*. MIT Press, Cambridge.
4. CHILD J (1972) Organizational structure, environment and performance: the role of strategic choice. *Sociology* **6**(1), 2–22.

5. COOLEY WW and LOHNES PR (1962) *Multivariate Procedures for the Behavioral Sciences*, Chapters IV and VI. Wiley, New York.
6. CYERT RM, DILL WR and MARCH JG (1970) The role of expectations in business decision making. In *Management Decision Making* (Eds. WELSCH LA and CYERT RM), p. 87. Penguin, Baltimore.
7. CYERT RM and MARCH JG (1963) *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, New Jersey.
8. KING WR and CLELAND DI (1974) Environmental information systems for strategic marketing planning. *J. Marketing* **38**, 35–40.
9. KING WR and CLELAND DI (1975) The design of management information systems: an information analysis approach. *Mgmt Sci.* **22**, 286–297.
10. KING WR and EPSTEIN BJ (1976) Assessing the value of information. *Mgmt Datamatics*.
11. KOTLER P (1966) A design for the firm's marketing nerve center. *Bus. Horizons* **9**, 63–74.
12. MURDICK RG (1970) MIS development procedures. *J. Systems Mgmt* (Dec.) 22–26.
13. NORMANN (1969) Some conclusions from 13 case studies of new product development. S.I.A.R. Report, Number UPM-RM-100, Stockholm.
14. PIPER JR, JR. (1971) How US firms evaluate foreign investment opportunities. *M.S.U. Bus. Topics*, **19** (3), 11–20.
15. RAO CR (1962) *Advanced Statistical Methods in Biometric Research*, Chapter VII. John Wiley, New York.
16. TAYLOR DW (1970) Decision making and problem solving. In *Management Decision Making* (Eds., WELSCH LA and CYERT RM). Penguin, Baltimore.
17. WEICK KE (1969) *The Social Psychology of Organizing*. Addison-Wesley, Reading, Mass.
18. WRAPP HE (1967) Good managers don't make policy decisions. *Harv. Bus. Rev.* (Sept.–Oct.)

ADDRESS FOR CORRESPONDENCE: *Professor WR King, 2125 Cathedral of Learning, Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

The Robustness of Linear Models for Decision-Making

ROBERT H ASHTON

(*Received March 1976; in revised form June 1976*)

This paper elaborates on some issues discussed by Moskowitz, who presented evidence that linear multiple regression models, estimated from decisions made by individuals, often outperform the individuals themselves. In discussing his results, Moskowitz (1) suggested that inconsistency in information utilization by individuals may account for the relative superiority of regression models, and (2) expressed concern over the robustness of linear regression models to changes in (a) information environments, (b) weighting parameters, and (c) functional form of the model. This paper discusses reasons (in addition to inconsistency) for the relative superiority of model over man, and it summarizes recent research in psychology concerning the robustness of linear regression models (and linear models in general). This paper is supportive, rather than critical, of Moskowitz's research.

Vol. 4
1976

MAN VERSUS LINEAR REGRESSION MODELS[1]

MOSKOWITZ [10] provided evidence that inconsistency in the utilization of information by man is a major factor in man's inferiority relative to models. This finding deserves additional elaboration not provided by Moskowitz. Note that Moskowitz was concerned with *linear* models. The extent to which a linear model can simulate an individual's decisions is reflected in the model's residuals. Linear model residuals may be composed of random error and/or nonlinear information utilization. In explaining why linear models of individuals often outperform the individuals themselves, it is useful to distinguish between two components of human decisions, systematic bias and random error, and between two possible modes of representing human information utilization, linear and nonlinear.

Decisions made by individuals reflect, in addition to valid information utilization, systematic bias and random error. Systematic bias occurs, for example, when the individual consistently overestimates or underestimates the importance of particular data items used in decision-making, while random error refers to *inconsistency* in the extent to which the individual relies upon

[1] Much of this section is taken from [1].

particular data items. While systematic bias is captured in regression models, the use of such models to make decisions eliminates the random error component.

Concerning the distinction between linear and nonlinear representations of human information utilization, several studies have shown that various types of nonlinear models (for example, curvilinear, configural, conjunctive, and disjunctive) provide very little, if any, increase in simulative power over that provided by simple linear models. Some problems of nonlinear representation discussed by Hoffman [8, p. 63] partially explain this consistent finding:

> ". . . an undirected search for configural relationships within a finite set of data is fraught with difficulties, and virtually doomed to disappointment. Without having literally tens of thousands of cases available, predictive schemes which randomly search for configurated terms can seldom be reliably replicated, for a very large number of such terms must be investigated, and the problem of parameter estimation becomes increasingly acute as the number of dimensions or terms is increased. Furthermore, the error in estimating the parameters of such models is adversely affected even more as the complexity of the individual terms is increased. This is so because the concept of configuration, implying, as it does, the *interdependence* of variables, demands the inclusion, in a general model, of terms which are functions of several variables, and which are therefore highly intercorrelated. Under such circumstances, the rare but significant configural parameter lies embedded in a sea of chance relationships, from which it cannot reliably be distinguished."

The importance of distinguishing between systematic bias and random error and between linear and nonlinear representations has been addressed by Goldberg [6, p. 424]: "To the extent that the (linear) model fails to capture valid nonlinear variance in the judge's decision process, it should perform less creditably than the judge. To the extent that it eliminates the random error component in human judgments, it should perform more validly than the judge." The evidence, e.g. [5, 7, 11, 14], implies that elimination of the random error component of human decisions is more important than is the failure of linear models to capture nonlinear information utilization.

Recent research in psychology [3] has revealed three additional reasons, not discussed by Moskowitz, for the relative superiority of linear regression models. First, in typical decision situations, the independent variables generally have, or may be rescaled to have, monotone relationships to the dependent variable. This results in a situation in which higher values of each independent variable are directly associated with higher values of the dependent variable, regardless of the values of the remaining variables. Second, the relative weights estimated from a linear regression analysis are not affected by error in the measurement of the dependent variable. Error reduces the expected values of all the weights by a constant. Third, error in the measurement of independent variables tends to make conditionally monotone functions *more* linear. Therefore, linear regression models work quite well because the contexts in which they are used tend to be conditionally monotone contexts in which there is sizable measurement error.

As noted earlier, Moskowitz expressed concern over the robustness of linear regression models to changes in information environments, weighting parameters, and functional form of the model. The above discussion indicates that linear regression models are robust over deviations from linearity (the functional form problem). Other research has indicated that linear models are robust with respect to weighting parameters and changed environments.

UNIT WEIGHTING MODELS

Weighting parameters

Moskowitz [10] asked "How robust are the linear regression models to differences in specification of the weighting parameters?" Recent research indicates that such models are extremely robust in this sense. For example, Dawes and Corrigan [3] found that linear models with weights chosen at *random* generally outperformed human beings. They constructed ten thousand models for each of five decision situations (which had been modeled previously in typical multiple regression fashion) by selecting normal deviates at random from a normal distribution with unit variance and using the absolute values of the deviates as weights for independent variables. Average validities indicated that these random linear models outperformed individuals in four of the five situations (the other was a tie). Further, random models outperformed individuals' regression models in three of the five situations, providing evidence that linear models are robust over deviations from optimal weighting.

Following Dawes and Corrigan, Einhorn and Hogarth [4] investigated the use of models with *unit weights*, i.e. where the independent variables are simply weighted by $-1{\cdot}0$ or $+1{\cdot}0$. Note that the use of unit weighting models requires only that one be able to specify (1) the independent variables to be included and (2) the sign of the zero-order correlation between each independent variable and the dependent variable. (If one is unable to specify even the sign of the relationship, he should probably exclude that variable from the model.)

Einhorn and Hogarth reasoned that unit weighting models should perform well (compared with models in which weights are estimated by regression procedures) because (1) sample size is unimportant since unit weights are not estimated from the data, (2) as a corollary to (1), it is possible to have virtually any number of independent variables in the model, (3) also as a corollary to (1), unit weights are 'estimated' without error, and (4) the stability of unit weights, unlike regression weights, is not influenced by intercorrelations among independent variables. Further, research in various other contexts, reviewed in [4], has shown unit weighting to be superior to various other types of weighting.

Analytical results produced by Einhorn and Hogarth, using 'expected mean squared error' as their measure of model fit, indicated that unit weighting models are not only robust, but will often outperform models whose weights

Vol. 4
1976

are estimated by regression analysis. A primary determinant of the relative efficacy of regression weighting and unit weighting was sample size. Deriving analytical results for the case of four independent variables, Einhorn and Hogarth found that only in the case of 'large' samples (defined by the researchers as $N > 200$) will regression weighting consistently outperform unit weighting. With 'moderate' sample sizes ($50 < N < 200$), their results were mixed but generally supported the use of unit weighting. With 'small' sample sizes ($N < 50$), unit weighting was found to be superior because regression weights were estimated with too much error. Further, a particular sample may be so small that regression analysis simply cannot be performed.

A second determinant of the relative efficacy of regression weighting and unit weighting was the degree to which the criterion (the dependent variable) is defined. Einhorn and Hogarth considered three possibilities: (1) a *well-defined* criterion is one which is (a) available, (b) measured with a moderate degree of reliability, and (c) related to an 'ultimate' criterion (in the sense, for example, that the criterion of 'grade point average' is related to the ultimate criterion of 'success in graduate school'); (2) a *vague* criterion is one which is available and measurable, but whose relationship to an ultimate criterion is questionable (in the sense, for example, that the criterion of 'salary' is a questionable indicant of 'success'); and (3) the *absence* of a criterion because of either conceptual or measurement problems. In their example with four independent variables, Einhorn and Hogarth found that for 'well-defined' criteria regression weighting was superior to unit weighting under two conditions. (1) large sample sizes ($N > 200$), and (2) moderate sample sizes ($50 < N < 200$) if $R^2 > 0{\cdot}50$; when $R^2 < 0{\cdot}50$, as is frequently the case in applied settings, unit weighting was indicated to be preferable to regression weighting. Preference for regression weighting or unit weighting in decision situations with 'vague' criteria depends on the degree of reliability and ultimate relevance of the criterion. Finally, in situations where criteria are absent, regression analysis cannot be performed.

Changed environments

Regression weighting may have additional disadvantages when compared to unit weighting across new samples. In a comprehensive review of the 'clinical versus statistical prediction' literature, Sawyer [12] suggested that

> "Predicting in a population different from the one in which experience was derived offers a possible relative advantage to (man because he, unlike model) can adjust to the altered situation. How successfully (man) adjusts depends, of course, upon his ability to recognize the differences in the two samples and to anticipate their effect on the relation between predictors and criterion."

Moskowitz's results, however, contradicted Sawyer's suggestion. Moskowitz [10] found that 15 of 21 individuals were outperformed by their respective linear regression models even when the models were used in an environment

different from the one in which they were derived. (Specifically, Moskowitz was concerned with a change in the forecast information received by subjects.) Moskowitz was reluctant to generalize regarding conditions under which 'inappropriate models' can be used effectively. However, given that most applications of regression models are to "samples from populations that differ somewhat from those on which the [regression weights] are estimated" [3], the work of Einhorn and Hogarth is again relevant. Unit weighting should perform well compared with regression weighting when one is extrapolating across samples. The reasons were discussed earlier: Unit weights are not estimated from the data and are, therefore, not sample-specific. Of course, the same independent variables must be relevant in the two (or more) samples, and the sign of the zero-order correlation between each independent variable and the criterion must be the same across samples.

Two further reasons, which lie outside the analytical results of Einhorn and Hogarth, that unit weighting models should be considered viable alternatives to regression models are: (1) whereas the theoretical work of Einhorn and Hogarth assumed that all assumptions of multiple regression had been met, this is probably rarely the case; therefore, their results were likely biased *toward* regression; (2) in a practical sense, the use of unit weighting should be considerably less costly than the use of regression weighting. For unit weighting, modeling of *people* is unnecessary.

Vol. 4
1976

EXTRAPOLATING FROM PSYCHOLOGY TO MANAGEMENT

Studies comparing the relative performance of man with that of linear models have dealt almost exclusively with decision problems in psychology, e.g. [6], or education, e.g. [13]. Of course, evidence generated in behavioral science disciplines cannot necessarily be extrapolated to other disciplines, such as management, for a number of reasons, including differences in decision contexts and conflicting results generated by behavioral scientists themselves. In this particular case, however, such problems do not appear to be as severe as in other areas. The literature comparing the performance of man and linear regression models of man is one of the most consistent literatures in psychology [11]. Further, it appears that the characteristics of decision situations in which models have been shown to outperform individuals (e.g. conditionally monotone relationships and error in the measurement of both independent and dependent variables) are pervasive [3]. In fact, Goldberg [6], in psychology, and Bowman [2], in management, reached virtually identical explanations concerning the relative superiority of models in their respective fields, although these two researchers were working independently. In discussing man, Goldberg [6] stated:

> "While he possesses his full share of human learning and hypothesis-generating skills, he lacks the machine's reliability. He 'has his days': boredom, fatigue, illness, situational and interpersonal distractions all plague him, with the result that his repeated judgments of the exact same stimulus configurations are not identical. He is subject to all those human frailties which lower the reliability of his judgments below unity. And, if the judge's reliability is less than unity, there must be error in his judgments—error which can serve no other purpose than to attenuate his accuracy. If we could remove some of this human unreliability by eliminating the random error in his judgments, we should thereby increase the validity of the resulting predictions."

Bowman [2] distinguished between systematic bias (which he called simply 'bias') and random error (which he called 'variance') and stated:

> "It seems useful to attempt an explanation of why decision rules derived from management's own average behavior might yield better results than the aggregate behavior itself. Man seems to respond to selective cues in his environment—particular things seem to catch his attention at times . . . while at other times it is a different set of stimuli. . . .
>
> Departures of the decision making behavior of management from the preferred results, in this sense then can be divided or factored into two components, one which in the manner of a grand average departing from some preferred figure, we call bias (which causes a relatively small criteria loss due to the dish shaped bottom of the criteria surface), and one which representing individual occurrences of experience departing from the grand average, we call variance (which causes larger criteria losses due to the individual occurrences up the sides of the criteria dish-shaped surface). It is the latter and more important component which seems to offer the tempting possibility of elimination through the use of decision rules incorporating coefficients derived from management's own recurrent behavior."

Given the similarity of (1) empirical results in psychology and management and (2) the explanations of Goldberg and Bowman, it is difficult to agree with Moskowitz's contention that Goldberg's formulation of the 'man *vs.* model' issue is inappropriate from the perspective of management [10, p. 679, n. 3]. Moskowitz stated that whereas the Goldberg-type studies are concerned with predictability, management researchers are concerned with monetary payoff functions. While psychologists have not completely ignored (monetary or non-monetary) payoff functions, e.g. [9, p. 347], this is an important distinction between the two disciplines. However, it is perhaps more meaningful to view the Goldberg-type studies as a necessary and important intermediate step toward the modeling of human decisions to which monetary payoff functions are sensitive. By viewing the psychological studies in this manner, researchers in management should decrease markedly the likelihood of overlooking important prior theoretical and empirical work.

REFERENCES

1. Ashton RH (1975) User prediction models in accounting: an alternative use. *Acct Rev.* **50,** 710–722.
2. Bowman EH (1963) Consistency and optimality in managerial decision making. *Mgmt Sci.* **9,** 310–321.

Vol.
197

3. DAWES RM and CORRIGAN B (1974) Linear models in decision making. *Psychol. Bull.* **81,** 95–106.
4. EINHORN HJ and HOGARTH RM (1975) Unit weighting schemes for decision making. *Orgn. Behav. Hum. Perf.* **13,** 171–192.
5. GOLDBERG LR (1968) Simple models or simple processes? Some research on clinical judgments. *Am. Psychol.* **23,** 483–496.
6. GOLDBERG LR (1970) Man versus model of man: a rationale, plus some evidence, for a method of improving on clinical inferences. *Psychol. Bull.* **73,** 422–432.
7. GOLDBERG LR (1971) Five models of clinical judgment: an empirical comparison between linear and nonlinear representations of the human inference process. *Orgn. Behav. Hum. Perf.* **6,** 458–479.
8. HOFFMAN PJ (1968) Cue-consistency and configurality in human judgment. In *Formal Representation of Human Judgment*. (Ed. KLEINMUNTZ B). Wiley, New York.
9. HOFFMAN PJ, SLOVIC P and RORER LG (1968) An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. *Psychol. Bull.* **69,** 338–349.
10. MOSKOWITZ H (1974) Regression models of behavior for managerial decision making. *Omega* **2,** 677–690.
11. RORER LG (1973) A circuitous route to bootstrapping selection procedures. *Oregon Res. Inst. Res. Bull.* **12**(9) 1–31.
12. SAWYER J (1966) Measurement *and* prediction, clinical *and* statistical. *Psychol. Bull.* **66,** 178–200.
13. WIGGINS N and KOHEN ES (1971) Man versus model of man revisited: the forecasting of graduate school success. *J. Pers. Soc. Psychol.* **19,** 100–106.
14. YNTEMA DB and TORGERSON WS (1961) Man-computer cooperation in decisions requiring common sense. *I.R.E. Trans. Hum. Factors Electron.* **2,** 20–26.

ADDRESS FOR CORRESPONDENCE: *Professor RH Ashton, Department of Accounting, The University of Texas at Austin, College of Business Administration, Austin, TX* 78712, *USA*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 5, 1976. Pergamon Press. Printed in Great Britain

Planning Ambulatory Health Care Delivery Systems

VF DÖKMECI

(*Received June 1975; in revised form April 1976*)

This paper develops an optimal ambulatory health care delivery system for a community, by determining the optimal number, size and location of health centers. The demand for ambulatory health care services is given. The objective is to maximize the difference between the benefits of the patients receiving treatment and the total cost of the system. Utilization of facilities, which is a function of distance, is expressed in terms of benefits and incorporated into the cost–benefit analysis. The model is illustrated by a numerical example.

Vol. 4
1976

INTRODUCTION

PLANNING of health facilities on the community level has long been of primary concern to health facility planners. Their purpose is the efficient utilization of scarce resources and personnel and a balanced distribution of facilities with respect to demand. However, decisions concerning the size, number and location of health care facilities in a community are rarely based upon an adequate method. Groups desiring to obtain public funds for the construction or expansion of facilities are required to submit their proposals to planning agencies for review, but the criteria used for acceptance or rejection are largely intuitive. A more elaborate discussion of the subject is given in [2, 7, 17]. The few attempts to develop useful methods for optimal distribution of health facilities [1, 14, 15] exclude the coordination among the different types of health care facilities.

The purpose of this paper is to develop a general model to determine the optimal number, size and location of health centers by taking into consideration the economies of scale and the coordination between the health centers and other facilities. By coordination is meant the movement of patients, personnel, and equipment between the health centers, intermediate hospitals and the regional center so as to provide comprehensive health care and efficiency in services. The regional center provides superspeciality care, research and

education; intermediate hospitals offer speciality care and general care, and research, and health centers primary and preventive care. The locations of patients, the regional center and intermediate hospitals, and demand for health centers are given. The objective is to maximize social net benefits, that is the difference between the social benefits of patients receiving treatment and the total cost (travel cost + facility cost) of health centers [10]. Although the travel cost is paid by the patients, it is still taken as a variable since we are dealing with social benefits. Utilization of health centers is taken to be a function of distance, expressed in terms of social benefits and incorporated into the cost–benefit analysis. Several alternative systems with different number, size and location are evaluated. The alternative with maximum net benefits is chosen as the optimal system.

UTILIZATION OF AMBULATORY HEALTH CARE

Previous studies [3, 12] have shown that distances between health care facilities and patients' residences considerably affect the utilization of the health services. There is a different distance decay function for each level of health care, the slopes of these functions being used as indicators of the distance effect. The smaller the slope exponent, the larger is the health care facility service area. The larger the exponent, the fewer is the number of people who would travel an appreciable distance. In addition, the exponent is related to medical speciality in a way that people are willing to travel further to specialists than to general practitioners. The expression of the distance decay function can be written in general terms as

$$u_{ij} = \frac{n_j}{d_0 + d_{ij}^f}$$

where u_{ij} is the number of patients going from the point j to the facility i, n_j the number of patients at the point j, d_0 a constant, d_{ij} the distance between the facility i and location j, and f the exponent. In this study, we shall take $d_o = 1$ and $f = 1$ for the health centers which only offer one level of medical care.

The total utilization of health centers is expressed as

$$U = \sum_{i=1}^{q} \sum_{j=1}^{p} \frac{n_j \gamma_{ij}}{1 + d_{ij}} \tag{1}$$

where p is the given number of demand points (patient locations, intermediate hospitals and the regional center), q the number of health centers, and γ_{ij} the elements of the allocation matrix. A demand point can only be allocated to one

health center; $\gamma_{ij} = 1$ if the health center i supplies the point j and $\gamma_{ij} = 0$ otherwise.

In this study, a general benefit function is taken into consideration without going into its detailed analysis. It reflects the decision-maker's expectations concerning the present value of future private benefits deriving from the utilization of the facilities by the community members. In order to be able to compare benefits with the cost of services, the benefits to each patient utilizing facilities must be converted to a dollar equivalent (for various suggestions see [10, 13, 16]). If the average benefits per person is k, the benefit function is simply Uk.

TRAVEL AND FACILITY COSTS

In several previous health facility location studies (for instance, [5, 11]), the transportation cost is taken as a component of the total social cost of the health facility systems. This is justified by Coughlin, Isard and Schneider's study [6] which showed that the transportation cost alone takes a significant proportion of the total cost of the hospital. By including time and other relevant costs, and the recent increase in transportation cost, the travel cost becomes an important component of the total social cost of the health facility system. By taking into consideration the interactions between the facilities and patient locations, and among the facilities, the analytical representation of the transportation cost is

$$T = \sum_{i=1}^{q} \sum_{j=1}^{p} u_i \gamma_{ij} d_{ij} ts \tag{2}$$

where t is the unit transportation cost and s the average trips per person.

The facility cost, H_i, is taken here as the sum of capital and operating costs, (based on [9]) i.e. $H_i = a_i + b_1 K_1$ $\left(\text{the total facility cost being } F = \sum_{i=1}^{q} H_i\right)$, where a_i and b_i are constant and K_i is the number of service units provided by the health center i namely

$$K_i = \sum_{j=1}^{P} u_i \gamma_{ij}.$$

THE MODEL AND SOLUTION

The model can now be stated: determine the number, size and location of health centers in order to maximize the net benefits to society of providing

Vol. 4
1976

health services subject to constraints of service requirements, namely find the locations of the centers and the number q to maximize

$$Z = Uk - (T + F). \tag{3}$$

Aside from some simple cases, the solution of this type of problem cannot be obtained in a single process. Therefore, a two-step heuristic procedure is applied. First, the optimal locations are determined with minimum distance criterion for varying number of health centers. Second, using the locational coordinates, the net benefits are computed for each alternative of different number of health centers. That is:

Step 1. Select the number of health centers. For consistency, this number should be equal or greater than the total number of intermediate hospitals and the regional center. Then, the demand points are decomposed into subsets, by using the 'Eliminate-alternate-correction heuristic' method (see, for instance, [4]). The method is based on minimum distance criterion, and it assigns a health center to a known subset of demand points. The effects of distance on utilization is excluded as a first approximation, and for each subset, the optimal location of health centers are determined by minimizing the sum of distances between the patient and health center locations, namely,

$$D = \sum_{i=1}^{q} \sum_{j=1}^{p} n_j \gamma_{ij} d_{ij}. \tag{4}$$

This means a quasi-dual problem will be solved in order to overcome the difficulties rising from the solution of equation (3). In order to minimize equation (4), the necessary conditions are

$$\frac{\partial D}{\partial x_i} = \sum_{j=1}^{p} \frac{n_j \gamma_{ij}(x_j - x_i)}{d_{ij}} = 0; \quad \frac{\partial D}{\partial y_i} = \sum_{j=1}^{p} \frac{n_j \gamma_{ij}(y_j - y_i)}{d_{ij}} = 0 \tag{5}$$

where x_i and y_i are the coordinates of center i. These conditions can be shown analytically to be sufficient to assure a minimum. As (5) may not have an explicit solution, an iterative procedure is applied [4, 8]. As a result, the optimal locations of health centers for the chosen number of facilities are obtained. The same procedure is repeated for varying number of health centers.

Step 2. The locational coordinates are substituted into (3), and the net benefits are obtained for each alternative system consisted of different number of health centers. The alternative which has the maximum net benefits is chosen as the best system among those considered.

APPLICATION

The model was applied in a selected area to determine the optimal number, size and location of health centers. The location of patients, intermediate hospitals, the regional center and their weights were prescribed. The different number of health centers was selected for each alternative (varying from two to twelve).

By using a computer program written in Fortran and run on the IBM 360/91, for the solution procedure described earlier we determined the locations of the health centers for each of the eleven alternatives. Then, we calculated the total utilization from equation (1) and the total cost from the other equations, (where $a_i = 250$ and $b_i = 22$).

According to the results, as the number of health centers increases, net benefit function increases until it reaches a maximum since both the increasing benefit function and decreasing total cost function have positive effect on net benefits. After the maximum point, the net benefits decrease because the negative effect of the increasing total cost is more important than the positive effect of the increasing benefit function.

Vol. 4
1976

CONCLUSION

This paper develops a heuristic planning model to assist the decision-maker to choose the optimal ambulatory health care delivery system among several alternatives with different number, size and location. A quasi-dual problem is solved and the optimal health center system is approximately determined with respect to the maximum net benefits criterion. The results from using the heuristic model should influence the way decisions are made simply by changing the informational context within which the decision making process takes place.

According to the numerical analysis, while economies of scale encourages the centralization of health centers, utilization factor encourages decentralization. We observe that public interest is in neither total centralization, nor total decentralization, but in a balanced state between two extremes.

It is worth noting that the general techniques and procedures developed in this study might also prove useful in the planning of other service facilities such as post offices, libraries and schools, in which demand is distributed over a geographic area.

ACKNOWLEDGEMENT

The author is indebted to Prof. Hande Suher for her encouragement and to the referees for a careful and constructive critique of this study.

REFERENCES

1. ABERNATHY WJ and HERSHEY JC (1972) A spatial-allocation model for regional health-services planning. *Ops Res.* **20**(3), 629–642.
2. ANDERSON R (1971) The public's view of the crisis in medical care: an impetus for changing delivery systems. *Econ. Bus. Bull.* **24**, 44–54.
3. CIOCCO A and ALTMAN I (1954) Medical service areas and distances travelled for physician care in Western Pennsylvania. US Publ. Hlth. Serv., Publ. Hlth. Monogr. No. 19.
4. COOPER L (1967) Solutions of generalized locational equilibrium models. *J. Reg. Sci.* **7**(1), 181–198.
5. COUGHLIN RE (1964) Hospital complex analysis: an approach to analysis for planning a metropolitan system of service facilities. PhD Dissertation, University of Pennsylvania, Philadelphia, PA.
6. COUGHLIN RE, ISARD W and SCHNEIDER JB (1964) The activity structure and transportation requirements of a major university hospital. Discussion Paper Series, No. 4, Regional Science Research Institute, Philadelphia, Mimeo.
7. DÖKMECI VF (1972) An analytic planning approach to regional health facility systems. PhD Dissertation, Columbia University, New York, NY.
8. DÖKMECI VF (1973) An optimization model for a hierarchical spatial system. *J. Reg. Sci.* **13**(3), 439–451.
9. FELDSTEIN P (1961) *An Experimental Investigation of the Marginal Cost of Hospital Services.* Chicago.
10. KLARMAN HE (1965) *The Economics of Health.* Columbia University Press, New York.
11. LONG MF and FELDSTEIN P (1967) Economics of hospital systems, peak loads and regional coordination. *Proc. Am. Econ. Ass.* 119–122.
12. MORRILL RL and EARICKSON R (1968) Hospital variation and patient travel distance. *Inquiry* **5**, 26–34.
13. RICE D (1966) *Estimating the Cost of Illness.* Health Economics Series, No. 6, US Publ. Hlth. Serv.
14. SCHNEIDER JB and SYMONS JG (1971) Regional health facility system planning: an access opportunity approach. Discussion Paper Series, No. 48, Regional Science Research Institute, Philadelphia, Mimeo.
15. SCHULTZ GP (1969) Facility patterns for a regional health care system. Discussion Paper Series, No. 34, Regional Science Research Institute, Philadelphia, Mimeo.
16. WEISBROD BA (1960) *Economics of Public Health: Measuring the Economic Impact of Diseases.* University of Pennsylvania Press, Philadelphia, PA.
17. WHEELER KE (1969) The regional health authority: a possible approach to rural health problems. *Inquiry* **8**, 15–31.

ADDRESS FOR CORRESPONDENCE: *Assist. Professor VF Dökmeci, Mimarlik Fakültesi, The Technical University of Istanbul, Taksim, Istanbul, Turkey.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

EDITORIAL

How to Charge for Management Services

IMAGINE a company with several operating divisions, each one engaged in a clearly defined activity, such as the manufacture of products, or the provision of services to members of the public or to other companies. If a given division is totally isolated from the rest of the enterprise, in that it gets material and other inputs from outside sources and supplies its outputs to outside agents, it is a relatively simple exercise to draw a balance sheet for the activities of that division. There are, of course, problems to be settled relating to the time span which the balance sheet should represent, to the allocation of activities to different time periods, to the value that should be assigned to unfinished work (even to finished but as yet unsold goods) and to purchased materials, to the manner in which capital goods should be written-off during the interval in question, in short to all the familiar problems for which accounting conventions are supposed to provide us with necessary guidelines in order to arrive at an honest and true picture of the division's assets and performance. These problems should in no way be minimized, and controversies will continue to rage in the accounting world about the efficacy or propriety of various practices, which can often produce entirely different pictures of the same business entity. But within a given set of conventions, the principal proposition is that the entity should be examined by the application of an input–output analysis, and the profit and loss account becomes a reflection of the difference between the value of the outputs (revenue) and the total value of the inputs (costs), adjusted as necessary by the change in the value of the system's inventory.

The crux of such an analysis is that—apart from having to make value judgements on the assets of the system—the values of the inputs and the outputs are determined by external price mechanisms. The management of the division in question does not unilaterally fix the price it is prepared to pay for its material and other inputs, nor can it (unless it enjoys a monopolistic advantage) charge whatever it likes for its outputs, though it may be free to negotiate and thereby affect the price mechanism, and it may then be able to decide whether to proceed with or suspend certain operations.

But when a division is linked with others in a way that makes the outputs of one the inputs of another, a new set of relationships often emerges with pro-

Vol. 4
1976

found implications on the resultant perceived performance of each division. This set of relationships is reflected by the concept of transfer pricing and by the degree of freedom allowed to each division to disengage from a linkage with another division, whether its role is that of a supplier or a consumer of goods and services from that division. When a particular division is completely free in this respect and may treat the others as if they were outside agents, it can proceed to compare alternative prices and terms and then make decisions which are most beneficial to its own well-being and disregard the possible detrimental effects on other constituent parts of the company.

Such an arrangement has the obvious attraction in that it prevents feather-bedding and puts the onus on the management of each entity in an enterprise to account properly for its stewardship and performance. The company then becomes a collection of independent entities, held together under the umbrella of a holding company which derives its synergistic advantages more from a central financial muscle than from operational directives. This is particularly true for conglomerates with diverse interests and activities, where each entity is simply regarded as an investment, to be nurtured or disposed of as circumstances and opportunities dictate.

In a company which is not a conglomerate, pressures soon begin to mount to increase the level of inter-dependence of the constituent parts. The degree to which a division wishes to maintain its own freedom of action and yet constrain the freedom of another is often a reflection of market forces and the balance between supply and demand. If division A has a wide choice of suppliers and freedom to negotiate prices it is bound to be reluctant to take its supplies from division B, even at a market price, but should sources of supply become short, it would be keen to seek a secure supplier and to use the argument of company affinity to extract a long-term commitment from B. Similarly, if B is to be involved in a major capital investment in new plant, it would want to be assured of a secure outlet for its product and would welcome a binding arrangement with A.

In addition, pressures mount from head office, to ensure that potential benefits from vertical integration can indeed be realized, that central purchasing can reduce costs, that advantages can be obtained from pooling marketing resources, information systems, and technical knowhow. In some companies the result is a mammoth bureaucratic central organization for the headquarters functions, in others a conscious attempt is made to keep the head office relatively small in a coordinating and controlling role and so allow a fair amount of decentralization of staff functions.

These pressures, and the resultant organization structure, are naturally affected by the basic problem of how to charge for management services. In one respect it is a problem in transfer pricing, in that services provided by head office to a division can be charged to the division in the same way that physical goods are when they are supplied by another division; it is therefore possible to envisage

a charging relationship between head office and divisions following the same format as the transfer pricing mechanism set to operate between divisions. The format would then extend to enable the degree of freedom which any particular division is allowed in accepting or rejecting another division as a supplier or as a customer at a given price.

This notion that management services may be regarded as one of many inputs to an operating entity, and hence that its cost should be accounted for in the same way as other inputs are, has led to the development of several formulae for charging for management services. These formulae vary in detail and complexity, but two main schools of thought may be discerned. The first believes that management services should be treated in the same way as all other head office expenses: if the latter are allocated to operating divisions as general overheads (for example, in proportion to revenue) then management services should form part of this allocation, and if head office expenses remain unallocated (in order to isolate items of expenditure which lie outside the control and responsibility of the operating divisions) then the same principle should apply to central management services. The second school of thought argues that a distinction should be made between management services and other head office functions and concludes that management services should be charged for even when other head office expenses are not allocated to the operating divisions.

Vol. 4
1976

It is this latter case that concerns us here, the case where management services are singled out for special treatment compared with other overheads. The logic underlying this approach is embedded in a set of assumptions such as the following:

- for an overall direction of an enterprise, its head office must be able to assemble and analyse information in certain areas (accounting, finance, manpower, sales), which are deemed essential for the control exercised by the chief executive and his colleagues, whereas management services are not considered vital for such a control function

- an effective way to ensure that management services remain primarily concerned with operations is to give the operational entities a say in determining the programme of work of management services units

- management services, as the name suggests, should be set up as a facility for providing a service to help solve problems encountered and perceived by operating men, and this can best be achieved in an environment in which management services are seen to be in demand and in which they enjoy the confidence of the operating divisions

- for the operating man to regard the management services man as his servant, and not as his controller, he should assume a managerial responsi-

bility for projects conducted on his behalf and for any implementation that may ensue, although responsibility for the technical conduct of projects must remain with the professional men

- there are clearly many advantages to be gained from having a central management services group (as opposed to having smaller units dispersed among the operating divisions), not least amongst these are the opportunity to build up technical expertise and the ability to provide career development for existing and new personnel

- since operating men cannot always be relied upon to initiate studies to evaluate the performance of their activities, either because of lack of appreciation of what such studies can achieve or because of vested interests, there is a need to provide every incentive and every stimulant to these executives to ensure that whatever potential good management services can do is indeed fully realized.

These are the premises, and it is not difficult to see that various charging policies and organization structures can emerge as a result, depending on what weight one assigns to each of the premises (and particularly in view of the fact that they are not always compatible with each other). The case for charging largely rests on the argument that if the operating divisions are regarded as profit and cost centres, and if management services are primarily concerned with operational matters, it is then appropriate to maintain the expenditure involved only when the services are clearly seen to be in demand. The true cost–benefit test of the services, it is further argued, can best be provided through a price mechanism: it is only when people are prepared to pay for a service that they clearly indicate their acknowledgment of its value, and it is by refusing to pay or buy the service that they can register disapproval. It therefore follows that the existence of management services can be justified, and the level of expenditure can be regulated, by the demand generated in the full knowledge of the price-tag involved.

The consequences of this cold logic, which regards management services as a commodity looking for a purchaser, must be first that the services should be charged at their true cost, secondly that an operating division should be free to employ outside consultants instead of the company's central facilities, and thirdly that management services should be allowed to tout for business outside the company.

These consequences are often quite unacceptable to top management, and for good reasons. It is difficult to justify a situation where external consultants are called in while equally competent and reliable internal facilities remain idle (even for a short while), or where the internal facilities are overloaded with external work to the extent that they cannot be properly utilized for the benefit

of the operating divisions. Some constraints on the 'free market model' need then to be imposed. The first natural constraint is to prevent management services from setting up consulting relationships with current or potential competitors, followed by a directive for a clear priority to be given to internal vs external work. Such constraints are further backed-up by an inducement to an operating division to prefer the internal facility to external consultants in the form of a financial incentive, so that only a fraction of the total cost or only the variable cost element is charged instead of the full cost (one ingenious formula that I have come across allows X free consulting days followed by a charge for excess days, the charge being below the market rate quoted by external consultants).

Does the notion of 'subsidized management services' completely negate the fundamental premise that only the market price of a commodity gives a true reflection of its value? This is a debatable point, and has indeed been a centre of controversy among economists and in political circles. Believers in the leavening powers of the free market-place oppose subsidies as a matter of principle, while their opponents see subsidies as a compromise, which on the one hand averts the numbing effect of too high a price while on the other keeps in check what could become unbridled demand when the commodity in question is available freely.

Vol. 4
1976

But when it comes to management services, the case for a subsidy rests on shaky grounds. The distortion of the price mechanism annuls the prime purpose of introducing a change in the first instance, since it ceases to determine the proper amount of services that an operating division should consume. The demand for these services would obviously be affected by the level of the subsidy, the lower the price the greater the demand. It therefore follows that with any subsidy the demand is likely to be greater than it would be if the full market price were charged; a new mechanism would then have to be instituted to curb the degree to which this demand should be satisfied, and this can be achieved by putting an upper limit on the resources available to management services. It would appear, therefore, that while charging the market price is undesirable for the managerial reasons already stated, a subsidized price system cannot provide adequate guidance for determining how many resources should be allocated to management services, so that further control procedures must then be instituted for the purpose, procedures which often rely on subjective judgement of performance, on evaluation of progress reports and on feedback information, in short procedures which are akin to those employed in judging the effectiveness of other head office and non-production functions.

And as these same procedures can be used if no charging takes place at all, we have to conclude that no persuasive case can be made for any charges (at the full rate or at a subsidized level) for management services when other head office overheads are not allocated to operating divisions. This conclusion is further strengthened if we question the initial premise that the sole justification for the

existence of management services lies in the level of demand generated by the operators. While no sensible management scientist would deny that studies of industrial operations are likely to be more fruitful if they are actively supported by the executives directly responsible for these operations, situations do often arise, as mentioned earlier, where the initiative for launching such studies does not or cannot be expected to come from these executives; a subsidized price for management services may then become totally ineffective or irrelevant in providing the necessary incentive for the operating men, and the initiative must then come from head office, even if it may smack of central interference and control. To suggest that under such circumstances an operating division should be required to pay for studies which it did not ask for and which it may even need convincing are necessary, could be tantamount to adding insult to injury; it would then be less likely that a charging scheme could foster an atmosphere of cooperation, which is so desirable for the success of such studies.

Thus, committed as I am to the proposition that when we wish to evaluate the financial performance of a given entity our best course is to treat it as a cost/profit centre, I am bound to express grave doubts about the advisability of treating central management services differently from other central overheads. The case for treating them alike with respect to cost allocation seems to be overwhelming. This raises of course, the important issue of how to organize the central management services facility, how to control its expenditure, how to evaluate its performance, how to ensure that it can build up the necessary capabilities and that these capabilities are effectively utilized. But that is another story.

SAMUEL EILON
Chief Editor

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Vol. 4
1976

FEEDBACK

In-Company Training—A Comment

PROFESSOR EILON'S ten possible uses of training (Editorial, *Omega*, Vol. 4, No. 2) sums up admirably the thinking behind much of our own internal training within the National Coal Board. As he suggests, most of the stated purposes are intimately associated with the internal environment and inevitably point towards 'in-Company' training. This is particularly true of the mining industry with its special environmental problems and specific needs. For example, many of our managers are required to possess statutory qualifications which can only be gained by training and working within the industry.

The National Coal Board can lay claim to a long and successful experience of internal training, having established its own Staff College some 21 years ago and indeed ran courses elsewhere long before then. Staff College courses, formulated in consultation with the senior management of the industry are many and varied ranging from Junior Management Courses to Special Management Courses for 'middle' management to the Advanced Management Course for senior managers with special purpose courses as for example Operational and Colliery Management, Communication Skills Courses etc. for particular requirements.

To ensure that the teaching is relevant the College is in the main staffed with successful practising managers from within the industry but we are very conscious of the very real dangers of introspection and insularity in our own management thinking and training. It is for this reason that we invite speakers from universities, business schools and the establishments to assist our own teaching staff, and why, particularly on the more senior course, we offer a number of places to students from the energy industries, including foreign students. This, we feel gives a new dimension to our internal management training and access to outside research and ideas. The stimulation of new ideas and innovation is particularly necessary in the mining industry where, because of the statutory qualification requirement and for other reasons, staff turnover tends to be unusually low. Again it is for this reason that we send certain of our managers on external courses. For although the emphasis of our management training is on internal training I entirely agree with Eilon's conclusion that there is a need for both.

The role of training in shaping personal attributes to the needs of the organisation is a particularly interesting observation in the context of the emphasis which we in the NCB place on the link between training and our formal Management Development Scheme. The latter, based on the principle of 'Management by Objectives', places emphasis on training both as a means of assisting the individual to achieve his immediate objective and also to develop his career in the long term. In preparing Management Development reports senior managers are encouraged to be specific regarding their training recommendations and to ensure that these are relevant and meaningful. For the most senior of our management courses there is an added safeguard in that recommended candidates are interviewed and finally selected by a panel under the chairmanship of a Board Member. In this way we endeavour to avoid the pitfall of sending people on courses without a specific purpose.

The benefits of management training, as Eilon suggests, are difficult to measure: they are perhaps much easier to recognise than define. In some cases the benefit is more knowledge, in others an increased self-awareness and confidence, and in yet other cases intellectual refreshment. Of one thing I am in no doubt: the overwhelming evidence in favour of sound 'in-Company' training with, of course, the right blend of external training appropriate to the organisation and

the situation. At the National Coal Board we pursue a vigorous 'in-Company' training programme specially tailored to our needs, together with strong support for external management courses to provide a wider perspective.

SIR DEREK EZRA

Chairman
National Coal Board
Hobart House
Grosvenor Place
London SW1X 7AE
UK

Responsibilities of the Management Scientist

THERE were many topics covered in the intriguing editorial, "Responsibilities of the Management Scientist" (*Omega*, Vol. 3, No. 5, 1975) written by Professor Eilon. I wish to address myself to some part of the primary themes, namely to whom does the management scientist owe allegiance? (I could not take seriously the issue of management science responsibility in the conduct of concentration camps. The ethical issues are outrageous. I could not deal with the questions of South African gold mines because I do not know enough—except to state that the truth of any situation will eventually determine results.)

In my view, all such issues get cleared up when the planning horizon is long enough. For management scientists, responsibilities are strictly organizational; not personal, individual or divisional. Long-term, overall organizational objectives are the management science field's critical frame of reference. Professionalism is its credo. Departmental, divisional, and other subsystem interests are necessarily regarded as supportive rather than conflicting with the overall organizational goals.

It can be argued that the identification of the overall organization is difficult, e.g. with respect to multinational corporations. But that problem is solvable in a reasonable way. Where some ambiguity occurs, good multiple solutions probably exist, each of which satisfies the professional criteria.

Integrity of purpose is an essential of professionalism. Managers seek the advice of their accountants and lawyers (and even their doctors) because they believe them to be competent and credible. An accountant giving favored treatment to a particular department violates the organizational trust. Similarly, a management scientist who knowingly proposes solutions that are beneficial for a specific department but counterproductive for the overall organization violates the key professional ethic.

Management science will not survive as a profession if false-preference transgressions (i.e. subsystem optimization/system suboptimization) are common. I can attribute the occurrence of such suboptimization to two different causes. The first is the flawed-methodology reason, whereby the management scientist unwittingly overlooks the larger system's interactions. The second is the politically-conscious decision to assist the subsystem manager at the expense of overall (long-term) organizational benefits. The first cause can and must be removed by increasing the systems awareness of management scientists. The second cause can only be eradicated by insisting upon the highest professional standards for all management scientists.

Long-term organizational interests are usually (always?) congruent with social criteria. The dimensions of social judgment are many. I will mention only three: product quality, environmental quality and productivity (ergo cost). If the management scientist develops a new sequencing algorithm which improves productivity, but increases the probability of defective product or deteriorated service, then a tradeoff model is required. This tradeoff analysis examines both organizational and societal cost/benefits. Thus, in the matrix below sector I (better/better) dominates all other choices.

		Long-term society cost/benefits	
		Worse	Better
Short term organization cost/benefits	Better	III	I
	Worse	IV	II

Operations in sector II can only be sustained

in the short-term. Operations in sectors III and IV foretell long-term organizational trouble. With respect to the new sequencing algorithm, improved warranties, decreased prices or other compensations for the consumer should be considered. The new sequencing algorithm is not acceptable without such systems thinking.

The question: how will long-term sales volume (or share of market) be affected? goes a long way toward representing social interests. An organization with many loyal customers will not survive in the long-run if it contaminates the environment. It will change (as in sector II) or be regulated out of business if the society is serious about the environmental criterion. In this, and many other ways, the similarity between long-term organization success and societal goal realization is remarkable.

Short-term organizational interests often are pursued by owner–managers. Consequently, it may be beneficial that owner–managers are disappearing. Various factors have decreased substantially the length of the owner–manager's planning horizon. The deterioration in the continuity of the family firm, brings into play 'milking' strategies (sector III) which are aimed at maximizing short-term returns at the expense of long-term survival and success.

Unfortunately, however, the disappearance of the owner–manager does not seem to alter the preference for birds in the hand rather than in the bush. Short-term organizational interests are frequently pursued by non-owner managers. This occurs because managers act as if they were owners. Thus, even within a single firm, competition between departments, based on short-term goals, is hardly unusual. Even top-management, accustomed to years of 'in fighting', perpetuates the role of the owner, who must make it *now*. The sociological reasons for managers acting as if they were owners are not hard to explain. It takes great managerial skill to operate in sector I. Much less creative talent is required for sector III operations and the consequences can often be deferred to a successor's future.

What role does management science professionalism play in all of this? If the organization welcomes management science inputs, then it is adaptable and survival-prone. In turn, management scientists must recognize that they are responsible for more than immediate and local results. They must be responsible auditors of viable futures.

MARTIN K STARR

Columbia University
Graduate School of Business
401 *Uris Hall*
New York, NY 10027
USA

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Human Relations in Industry

MARCUS SIEFF[1]

(*Received March 1976; in revised form August 1976*)

In developed societies, good human relations become of crucial importance as people demand treatment as individuals. However, many in management positions pay only lip service to the implications of this need. The main reasons for this are that good human relations are very expensive in terms of both money and management effort and require fundamental revision of many inherited attitudes. They require appropriate organizational provision, good working conditions and amenities together with broad-ranging welfare facilities. Perhaps more important than these, however, are the intangible factors of caring and involvement by management and workers. Such principles have reached a high and successful stage of development in the author's company. Examples from several companies highlight the value of such hygiene conditions in diminishing disputes and facilitating essential change.

Leadership is crucial both in management and trades unions but effective leadership is only possible against a background of good human relations.

Vol. 4
1976

STANDARDS of education, literacy and knowledge for the majority in the developed world have improved considerably in the past 50 years. This has led many people rightly to expect a better standard of living and quality of life. They wish to be treated as responsible individuals who merit respect. Most want to be kept informed of developments which affect them; they appreciate their advice being sought. They long to feel wanted and many feel they have a constructive contribution to make. This applies particularly in the industrial and commercial fields, which for the sake of simplicity I shall refer to as industry. It emphasises the vital importance of establishing good human relations in industry. I use the term 'good human relations in industry' rather than the common term 'industrial relations' because we are human beings at work, not industrial beings.

Such relations cannot be successfully imposed on industry from outside, by government for example, though the government tries to do so from time to time. It is generally unsuccessful, witness in the United Kingdom the fate of Barbara Castle's "In Place of Strife" and Robert Carr's "Industrial Relations Act". State compulsion is generally a manifestation of desperation. Good relations can only develop organically from within the individual organisation.

[1] The Hon. Sir Marcus Sieff, Chairman, Marks & Spencer Limited.
This paper was originally presented as the Royal Institution Discourse, in London.

Of course nearly everybody in top management says that they believe in good human relations. In fact, I have found that management can be divided into three groups—those who pay lip service to the idea but do not care about it, those who believe in it but do not know how to implement it, and the small number who believe in it and implement the policy of good human relations successfully. Most top management does not appreciate what such a policy implies, certainly not what a total commitment to it entails.

Today, as organisations become larger and numbers employed increase it is apparently difficult to prevent personal relationships between senior management and employees from becoming remote; individuality is lost in numbers, people feel that they are tiny unimportant cogs in a vast machine and no-one is interested in their problems. They are often motivated by insecurity and sometimes by frustration and boredom. They become easy prey to the influential, dynamic, vociferous, militant minority who exacerbate differences and exaggerate grievances with the aim of changing the whole political and socio-economic structure. I do not refer here to the many trade union leaders and officials who genuinely want to solve problems constructively within the present system.

When management treats employees as responsible people meriting respect, most respond. This results in people working well, less absenteeism, greater staff stability, a more experienced staff and, generally, a willing acceptance of new and more modern methods of production and operation. In such circumstances, industrial strife is greatly reduced and sometimes eliminated.

Under such conditions, most people take pride in their work and this often results in high productivity and good profits. This in turn enables management to pay higher wages and provide those facilities which make for a contented and hard working staff, to have funds for investment which creates employment, and to pay dividends to the shareholders who have provided the money to operate the business.

My personal experience naturally colours my views. I have for the last 40 years been active in a business which believes in and implements a policy of good human relations. The Marks and Spencer philosophy was developed by the late Lord Marks and my father. The principles from which this philosophy emerged were laid down over 40 years ago. Perhaps the most important of these was to foster good human relations with all staff, with our manufacturers and suppliers, and with our customers.

In 1933 at the inaugural meeting of PEP (Political and Economic Planning), a non-partisan agency for economic and social research, my father said:

> "Lying at the core of our industry problem is a need for a new attitude of mind on the part of the human factors engaged in it. To break down the element of fear in human relations is difficult. If it can be done, a new spirit of co-operation will be born. There is certainly no short-cut to the achievement of this end. Yet without it, there can be little hope for the future of industry."

Depressing as it may seem, these words are as apposite today as they were 43

years ago. Remember they were spoken in the aftermath of the depression of 1929, with unemployment in excess of 2,500,000.

At that time, social security was so limited that some did not have enough to eat. Those who had jobs often shared their pay with the members of their family who were unemployed. In our staff restaurants we were shocked to find that some people were very hungry indeed. We became aware of personal unhappiness, insecurity, and frustration about the most elementary human needs, so we tried to provide not only for the material needs of our employees but also some kind of answer to the problem of alienation from the society which failed to provide these for so many.

Some remember, others have heard tales of the hardships of those days, and this consciously or subconsciously colours the attitude even today of many and, although, in these days of unfortunately increasing unemployment, the level of social security payments much reduces material hardships, it does not compensate for the indignity and frustration of the majority who genuinely seek work.

It requires more than words to implement a policy of good human relations. This has to start with top management having the right mental attitude which must be based on a sincere respect for the individual. Top management must understand the need for a right and just personnel policy, which is known to all employees and seen to be honoured in action. This must include a progressive wage policy so that all should share in the progress and prosperity of the organisation for which they work.

Now I would not like to delude you. The implementation of a good human relations policy requires sustained physical and mental effort, and it costs money. As organisations grow larger, it is impossible for those at the top to know everybody individually. So there must be a professionally trained group of people, whom we call staff management, to carry out the principles of personnel policy laid down by the Board of Directors, of whom a senior member's sole or main concern should be the implementation of an agreed personnel policy. But all executives should be aware of and interested in the personnel policy of their organisation.

The numbers in staff management should be sufficient to cover every employee individually and all employees should know with whom they can talk and discuss their problems. Top management must be in regular contact with staff management to ensure that they understand and are implementing the personnel policy laid down and, equally, are bringing to top management the views of employees.

But this is not enough; those in charge must visit the factory, shop floor and the locations where menial work is done with seeing eyes and listening ears. By seeing eyes and listening ears, I mean looking objectively at what is happening and listening objectively to what people have to say—not seeing what one wants to see or translating what people are saying into what one would like people to say. Then those in charge will learn a great deal about what is going

on, see what is being well done and should perhaps be intensified and, much more important, see what is being badly done and needs rectifying.

Top management cannot know how good or how bad staff amenities are unless they make use of them themselves, eat in the staff canteens or restaurants, see whether the food is decent and well cooked, whether the place is clean—visit the staff cloakrooms, lavatories and such other amenities as are provided. If amenities are not good enough for those in charge, they are not good enough for any employee.

Even today, when new factories are being planned, often too little attention is given to the provision of decent conditions for those who will work there and, even when provided, management often has little idea of how to use such amenities properly. Let me give you an example.

Some time ago I visited a large factory which had built a 'first class' restaurant for its thousand operatives. Although the surroundings were good and the meals generously subsidised, very few of the employees ate in the staff restaurant. I asked the Managing Director why. He said that he did not know, but then one of his assistants said in an aside to me: "the answer is quite simple, the facilities may be good but the food is bloody awful—but how should the top management know as they have never eaten in the employees' restaurant; their food is specially prepared."

I would like to give you some examples of how we operate in Marks and Spencer. In our Personnel Department we employ some 700 people trained in staff management whose primary responsibility is the well being of some 40,000 staff; that is approximately one for sixty. It is their job to see that the Board's policy is implemented. It is their job to see that every member of the staff is personally known and has regular contact with somebody who is concerned with his or her well being and progress. He or she is treated as an individual and with respect.

In every store, we have a staff management team which is part of operational management. Staff management know their staff well and have been trained, and integrated into the commercial life of the store. Therefore, they can train others from personal experience. A well planned training scheme for all grades of staff plays an important part in the development of good human relations. Apart from their operational role, staff management should be quick to sense when something is going wrong. Much of their work, like all good personnel work, is preventive.

The United Kingdom has a National Health Service, but despite this, there is much scope for individual medical care. Since 1933 we have provided advisory medical services and, since 1934, dental services. We encourage staff to attend regularly for examination and over 90 per cent do.

The majority of our staff are women, who spend much of their working day on their feet, so the setting up of a chiropody service—a small thing in itself—relieves discomfort for many and is much appreciated.

Again, the establishment of a system of testing for cervical and, recently, breast cancer has saved a number of lives and revealed many other ailments which, for one reason or another, the person concerned thought could not be cured, or just did not bother about. Among the first 40,000 examinations, we found seventy-five cases of cancer, of whom seventy-four are alive and well and seventy cases of suspected cancer now cleared, all of them followed up even if they have left our employment.

Some of these welfare activities are small things in themselves but they show people that management cares for them and is trying to study their personal problems and they, in turn, co-operate at work. The word 'welfare' has an old-fashioned sound reminiscent of the Victorian era, but I do not know a better one to replace it. People do have troubles and it is a fundamental part of a good staff policy to be able unobtrusively, and above all speedily, to give help and advice where needed.

The essence of any good welfare scheme is prevention rather than cure. A good system requires a minimum of bureaucratic fuss. I wish that more attention was given to the preventive aspect in the industrial scene where bureaucratic delays in solving minor problems and grievances end all too often in major industrial strife. A symptom of what I call the 'bureaucratic disease' is the increasing amount of documentation believed necessary to deal with people. In the middle fifties when we began to look at our paper systems, we carried out an operation called 'Good Housekeeping' and eliminated 60%. We cut out some 26,000,000 pieces of paper produced annually.

Legal requirements dictate a number of staff regulations, but the remainder should be reduced to a minimum and guidelines given to management to act sensibly and generously in dealing with people. If staff policy is well formulated and well understood, the manager on the spot can generally be trusted to make pretty sensible decisions within a broad framework. We tell our local store management "when dealing with human beings, if you are going to make a mistake, err on the side of generosity". This instruction is rarely abused.

There are also problems which cannot be dealt with locally or regionally and need a central decision. With us such problems are referred to the Welfare Committee which has met weekly without a break since 1934, over two thousand times. On average, eight cases are referred to the Committee weekly, advice being needed on how to deal with complicated individual problems or asking for permission to continue a certain course of action for which local authority has been exhausted. The Committee gives an immediate decision in 90 per cent of the cases and perhaps makes further enquiries in ten—but the point I want to make is that all cases are dealt with speedily and the individual concerned is informed of the decision or what further information is required within a few days.

Our welfare activites, which cover catering subsidies, medical, dental, chiropody and non-contributory pension scheme and other amenities, cost some £19,000,000 last year.

Vol. 4
1976

We discuss our personnel policies openly with other managements, some of whom think we must be mad to spend, as a Board, so much time on personnel work, to have so many people engaged in it and to spend so much money on it. I agree that perhaps we are mad but, looked at practically, our annual profit and loss account generally does not look too bad. In fact, those who lead Marks and Spencer have for a long time been convinced that without this investment in the care of people, we would not enjoy such success as we do.

Our staff know they are appreciated, that we have confidence in them, respect them as individuals and are determined to look after their welfare. They know they share in the progress and prosperity of the business. The great majority respond well, take pride in their work, their productivity is high, staff turnover and absenteeism low.

How many or perhaps how few of our large industrial concerns can say this and how many or how few people do they employ in implementing a policy of good human relations? One must sympathise with their problems. Some of the great firms have grown quickly, often by mergers and takeovers. The well established human relations of the original firm are sometimes overwhelmed by the very size of the new undertaking. Sometimes bad traditional practices of one of the merged firms become the standard for all.

In such circumstances, troublemakers thrive. Constructive moves to increase productivity, through investment or improved organisation, are misrepresented by those who want to subvert our socio–economic system into class-warfare between 'us and them'. Here a dynamic policy of developing good human relations is all the more important, but to get results requires time, patience and effort and must be a major and active plank of policy from the beginning.

The British Leyland Motor Corporation is an example of a company whose failure so far is in no small measure due to its inability to develop a spirit of co-operation between management and employees. So far as I remember, the original Morris management, led by William Morris, better known as Lord Nuffield, cared for people and relations were good. The Austin management did not have such good relations with their employees. When the two companies merged in 1952 to become the British Motor Corporation, problems due to poor relationship between management and employees arose.

When the successful Leyland Motors took over the leadership in 1968 did the new management of BLMC get their priorities right? Leyland had good relations with its employees but, to quote Geoffrey Robinson, former head of Jaguar Motors, part of BLMC, and who had previously held senior executive responsibility in BLMC:

> "The plain fact is that that essential precondition for success—the complete co-operation of the work-force, so rightly diagnosed in the first and increasingly, almost sententiously, emphasised in every successive chairman's statement—was not achieved. The merger was completed against a background of mutual suspicion between managers and men, of a plethora of piece work agreements, of overmanning, and of a minefield of jealously guarded

differentials. The full extent of the difficulties that lay ahead could hardly have been appreciated and it is hardly surprising that in the early period more awareness was shown of the potential than the problems."

He went on:

"Positive changes in industrial relations that were necessary within a relatively short period of time, if British Leyland was to survive, would have needed more vision and sheer self-confidence than the battle-weary line management of British Leyland could muster. But if the lead from management was along conventional and inadequate lines, so was the response of the unions."

Without personnel leadership on either side, mutual confidence was lost and resulted in what Mr Robinson describes as a 'conspiracy of disbelief' which went into a vicious circle. In October and November 1974, British Leyland lost 37% of its scheduled production through industrial disputes, quality faults, facility failures, absenteeism and so on. The company was bailed out by the Government, but the company has yet to prove that it can operate successfully.

People may say it is easy for a chain store organisation with between a hundred and two hundred staff in an average store, only 5000 men and 35,000 women not unionised, to carry out such a policy as ours. This is true, but Jaguar Motors managed to achieve in 1973/4, as part of British Leyland, continuous production through good industrial relations.

Some employees become bored because they have a purely repetitive job. This can and does lead to low productivity. Many people have a desire, possibly in some only sub-conscious, for creative expression. This is difficult to fulfil, as you know, on some modern production lines. We have paid too much attention to machines, too little to people and systems. No doubt advanced mechanisation and automation will eventually eliminate many repetitive tasks but some jobs will always depend on the human eye, human touch and human resourcefulness: progress starts with people who think. Machines do not think. Let me give you an example.

United Biscuits operate 12 factories in this country with 20,000 factory employees, 50% of whom are men, fully unionised. Production is highly mechanised and automated. There is much bored work. Management found it was impossible to make the work *per se* less boring, so it created a costly factory radio system which broadcasts on a 24-hour basis with news every half-hour, music, together with constructive and educational talks. There are a large number of individual loudspeakers so people can control the volume of sound in the area where they work or not listen at all if they do not want to.

This approach is now extended with a supply of portable cassettes to enable individuals at work to select language or other educational programmes, serial stories and classical music. All this has much improved the working environ-

ment, has reduced labour turnover and absenteeism and has markedly helped to increase productivity.

United Biscuits have deliberately set out to involve people in matters which concern them, and most behave responsibly. For example, stewards, as well as full-time union officials, take part in all their negotiations. Full disclosure is made to the unions of company results and objectives. Industrial leaders must educate. The gulf must be bridged between what management know, because they have access to a lot of information, and what staff believe, often in ignorance. There has been virtually no industrial strife at United Biscuits for many years.

You may be interested if I describe to you the rationalisation of United Biscuits and compare this with the attempted rationalisation in Upper Clyde. They are good examples of co-operation and confrontation. Upper Clyde Shipbuilders, you may remember, had a good deal of money pumped into it although it was really an amalgamation of four shipbuilding yards, three of which were not viable. When the Conservative Government came to power, they said they were not going to help lame ducks. They announced there would be no more money for Upper Clyde Shipbuilders.

Immediately the unions went on to the attack and said they would not accept this Government decision. In a very short space of time we had at the Upper Clyde one of the first of the famous sit-in strikes. There were 'work-ins' in all four yards throughout the Autumn. The company went into liquidation, part was sold off, staff was run down and the Government eventually had to give in and pump more money into the new Government-backed successor company. How different the result might have been if they had followed the example of United Biscuits.

United Biscuits grew during the 15 year period from 1953 to 1968, organically, by merger and by takeover. They found that four of their factories then operating in the United Kingdom, should be closed. From a purely economic point of view, they should have been closed within a few weeks, but instead United Biscuits discussed with the unions their plan for rationalisation, explained the position to the unions and said:

> "These factories are really redundant in the present set-up. They should be closed within a short period of time, but we realise that this means problems in the areas where these factories are, especially two that are in areas of high unemployment. We, therefore, propose to close these factories over the next 3 years, and see what we can do with redeployment, making sure people can get jobs in new and developing industries."

They also gave generous redundancy pay. They asked the unions for co-operation and received complete co-operation. They actually closed the factories over a period of 5 years and never lost one day's production because of it. On the final night of the factory closure, the Board of United Biscuits

Vol.
197

went down, gave a party for all the staff, and received only credit for the manner in which they had carried out the rationalisation.

Just think how different the position of Upper Clyde Shipbuilders might have been, and what friction and industrial unrest might have been avoided if the Government had said:

> "Upper Clyde Shipbuilders is not a viable company. It must be reorganised. From a purely economic point of view this ought to be done immediately, but we realise that this is going to cause industrial distress and unemployment, particularly in an area with already high unemployment. Therefore, we propose to streamline the operation over 3 or 4 years. It will cost a lot of taxpayers' money, but we know that the taxpayer will be willing to pay for this to avoid social distress. However, in order to carry this out we need the complete co-operation of the unions."

Then you might have had a very different picture. But perhaps that is the difference between well run industry which is run by the private enterprise and industry backed or run by the Government. Would it be different with outright nationalisation? I do not think so.

I have concentrated on the importance of giving priority to the building up of good human relations and thus developing a spirit of co-operation: but I would be quite wrong to claim that this is a cure-all for the problems which industry faces. I believe that the development of good human relations is very important but it alone will not solve industrial problems.

Let me touch briefly on some of these other problems, the solutions of which are of great importance. Some may consider they are as important or even more important than the development of good human relations. I do not.

I have spoken about inadequate leadership and lack of understanding of people's aspirations. In many cases it would be wrong to assume that material rewards are unimportant incentives towards individual effort and initiative, dedication and hard work, in fact to the development of leadership. Most people feel that, if they show initiative, assume greater responsibility and work harder, they should be adequately rewarded. Rightly or wrongly many people feel that the reward they receive today, for assuming greater responsibility and working harder, is not adequate compensation for the extra time and effort they are putting in. This goes right across the board—from top executives to works foremen.

Many of us are beset with increasing domestic problems brought about by inflation and high personal taxation which starts at a relatively low level, and many wonder whether the game is worth the candle. Some opt for TV and the quiet life. Among them are those who in the past would in all probability have eventually assumed considerable responsibility, acted dynamically and constructively. There are of course exceptions, but I am increasingly concerned at the diminishing number of capable people who wish to be leaders. There is an urgent need to reduce the high levels of taxation so that personal initiative and hard work are adequately rewarded at all levels.

Vol. 4
1976

Another major factor in the industrial scene is the power of the trade unions. In industry it requires two sides to co-operate and equally to confront. Trade union leadership falls more and more into the hands of those who do not believe that the solution of our problems lies in co-operation within our present socio–economic structure, but that the whole structure is wrong. They seek a system which would in effect be a state controlled, bureaucratically run economy with power residing in the hands of the trade unions, in fact in their hands. For many the real incentive is the acquisition of this power.

This group is a dynamic and often militant minority who in many cases do not represent adequately the views of the moderate majority of the members of the unions they claim to lead. They are elected by a very small minority of union members because the present system of election in many trade unions does not encourage the majority to exercise their democratic right to vote.

Lord George-Brown, writing on what he calls "the capture of critical positions of power by this"—as he also believes—"unrepresentative group of key figures in particular unions", considers that a compulsory, secret and preferably postal ballot would "quickly put to rights this problem of leadership which controls the use of trade union power". He is convinced that his fellow trade unionists would not merely acquiesce but positively applaud such urgently needed legislation. I agree.

During the last year, support has grown for moderate and constructive leaders, including those union leaders who seek solutions within our present socio–economic system in which the free enterprise sector of the economy plays a vital part. More union members are voting. Part of this improvement in union leadership is due to a greater appreciation that militant leadership offers no solution for our economic problems.

I have touched very briefly on these two subjects, trade union leadership and material incentives, and I do not minimise their importance. There are, also, other problems to which we must find solutions, but we shall not solve them unless we first learn to co-operate better in many areas of human activities.

To tackle today's problems we need firm leadership and sensible discipline; the best form of discipline is self-discipline; you cannot have good human relations without discipline. You cannot have discipline without good human relations. Such relations make cooperation possible—the alternative is conflict.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Cash Flow Models: A Review

GEOFFREY GREGORY

(*Received May 1976; in revised form June 1976*)

Cash flow management is concerned with the efficient use of the company's cash and short-term investments. Under stable economic conditions, this is a matter of deciding when to transfer assets and how much. The development of models designed to answer these questions is reviewed. More recently, emphasis has been placed on decisions concerned with operating conditions with the object of maintaining a level of cash flow in accordance with unstable conditions, in particular with a high rate of inflation.

Vol. 4
1976

1. INTRODUCTION

THE FLOW of money within an industrial community is simply the means whereby the various cells (companies, consumers) exchange goods, services and labour. The inter-dependence of the cells is characterised by this money flow, and a healthy flow is symptomatic of a healthy industrial society, at least under a basically capitalist system. The complete system can be depicted as a network of flows, where nodes are the companies or consumers, the actual flows being dictated partly by the state of the system and partly by the management decisions taken within each cell. Furthermore, some of the management decisions are concerned with the cash flow itself, and it is this aspect of the system which forms the topic of the review.

The benefits of good management of cash flow within a firm have long been apparent. Recently attention has been drawn to the perils of cash flow deficiencies, largely because the effects of a high rate of inflation rapidly become apparent as a company runs out of money for its day-to-day commitments. There has also been, as for example Sizer [27] pointed out, a serious effect not only on working capital but on the total investment that a company could make.

In cash flow management we are concerned with the balance between what the company keeps in cash—that is, in a low or zero interest bearing form, but readily available—and what has been placed in short term investments such as Treasury Bills, Commercial Bills, Deposit Accounts etc. The latter category, of course, brings in a significant interest rate and can be converted to the former fairly readily, usually in no longer than a week. Transfers in both directions cost money.

Most of the published work on cash flow models has been concerned with transfer policies. At the end of this review suggestions are made on how the models might be extended to answer some of the more immediate problems of cash flow management.

2. MODELS, DEFINITIONS AND NOTATION

A number of features were common to all models considered and these have been listed for convenience of reference in this section. Costs have been defined without specifying the units.

All flows of cash into and out of the company were termed *transactions*. These were payments of wages, taxation, bills for raw materials etc., and receipts of payments from customers.

The company kept its cash in what was termed a *two-asset environment* (this aspect of the model was subsequently extended). Readily available money was held in the *cash balance*, and short term investments in the *portfolio of short term liquid assets*. Interchanges of the assets were termed *transfers*.

Then

ν = daily rate of return from the portfolio of liquid assets. ν is also taken as the opportunity cost of holding the cash balance;

$\gamma(j)$ = cost of transfer of an amount j, assumed the same for either direction.

A desirable lower limit on the cash balance was assumed. In practice this would never be below zero, since an overdraft would obviously be more costly than the loss of earnings through realising some of the short term assets. It may, in fact, be higher than zero since banks may require a positive minimum or average balance in lieu of their charges. The limit was called the *compensating balance requirement*.

The above is the basic terminology and notation. Where a particular model required terms relevant it to alone, these terms were defined with the model.

The objective in all the models described was that of minimising the average cost per day of managing the cash balance. The analysis was very similar to that of the standard inventory control theory, even to the extent of deriving the simplest deterministic optimum in the familiar square-root form.

3. BAUMOL'S MODEL

The first attempt to fit a mathematical model to the cash flow problem has been attributed to Baumol [2]. Baumol's model, published in 1952, was a direct adaptation of the (s, S) inventory policy. It assumed only cash outflows, and the stock of cash was replaced periodically by the sale of assets of value M. Moreover, the cash outflows were assumed to be at a constant rate c per day.

Thus, since the whole model was deterministic, clearly the optimum policy was to run the cash balance down to zero (or whatever was the compensating balance requirement) every $c^{-1}M$ days. The graph of cash balance against time had the saw-tooth appearance and the only problem was the optimisation of M. Using standard calculus minimisation methods, the optimising value of M,

$$M^* = \sqrt{\left(\frac{2\gamma c}{\nu}\right)}.$$

A variant on the Baumol model was subsequently proposed by Sastry[26]. This allowed the cash balance to become negative but charged an amount u per money unit per day of deficit. Again the parallel with stock control was present. Sastry's optimum cash replacement became

$$M_0^* = M^* \sqrt{\left(\frac{u}{\nu+u}\right)}.$$

The same result was also derived by Lockyer [18] who went on to consider the case where the deficit situation had to be covered by a fixed-magnitude, fixed-time loan.

Clearly, there were two serious drawbacks to Baumol's model. It referred only to one-way cash transactions and transfers and it assumed a constant and deterministic environment. In times of economic stability (no references known) the defects might be conceptually overcome since a steady production output would result in a steady inflow of cash which, in turn, could be converted to short-term assets to be tapped at regular intervals. More practically, the model could only give a first approximation—albeit readily derived—to a policy.

4. PATINKIN'S MODEL

Some of the objections to the Baumol model were removed by the model proposed in 1956 by Patinkin [22]. Patinkin introduced the idea of cash flow transactions in the two directions of payments and receipts. He pointed out that in considering monetary demand for a specific period (e.g. for the next week) receipts and payments would not necessarily occur in a systematic way, so that an exact *pro rata* cash balance could not be kept over the week. In the case where the total cash flows each way were known for the week, it was therefore not sufficient to argue that the final total deficit or surplus was all that need be considered. In particular, if the majority of the payments had to be made before any income was received, the cash balance could fall below zero and costs thereby incurred. Patinkin's basic argument was that a company held a cash balance sufficient to cover all likely variations in cumulative transactions up to the time when the next decision could be taken to transfer short term assets into

cash (or vice versa). Too much emphasis on the subsequent mathematical analysis has obscured the value of this argument, and it is clearly incorrect [21, p. 49] to attribute to Patinkin only the special case considered in the analysis. This analysis (attributed to Dvoretzky) assumed that during any interval between transfer decisions there would be $2N$ transactions, half payments, half receipts and each involving the same amount, a, of cash. The transactions occurred in random order within the interval. If the initial cash balance was M, the value of the cash balance at any time during the week was a state of the simple random walk with states $M + Na$, $M + (N-1)a$,. . ., $M-Na$. Dvoretzky's analysis optimised the management cost, this being the sum of the expected opportunity cost and a penalty cost incurred when the cash balance fell below zero.

Again there are criticisms which can be made of the model. Cash outflows vary considerably in their size and incidence, from regular and frequent payments for say wages and salaries to the annual payment of tax. Generally the inflows are less frequent, but representing payments by customers they tend to be steady in value. In the sense that major and infrequent payments such as taxation can be planned and will be paid out of short term assets, these can be omitted from the analysis. (All models assumed that the amount in short term assets is effectively infinite, achieving this by a further transfer from long term assets.) With this modification the Patinkin model does then appear as a real advance on that of Baumol. Clearly, it provided a good insight into the basic cash flow problem.

5. THE MILLER–ORR MODEL

The Miller–Orr model [19, 21] incorporated many of the features of Patinkin's model. It was basically a two-asset model with cash flow transactions in both directions, but the Patinkin restriction of a finite time horizon was removed. Transfers between the two assets could occur at any time, and effectively the controlling variable in the management policy was the time at which transfers were made. With cash inputs and outputs occurring in random fashion, the cash balance was constrained to vary between two limits. On reaching either limit a transfer was made which returned the cash balance to some specified point within the limits. As with the Patinkin model, the cash balance was a random walk, but here there was no constraint requiring the cash balance to adopt some preassigned values at specified points in time.

Specifically, the model assumed that cash flows were generated by a stationary Gaussian random walk. Taking the working day as a convenient time unit, cash transactions occurred t times per day, where t^{-1} was some small fraction of the day. Miller and Orr suggested 8 as an appropriate value of t, implying hourly transactions, but the choice of t did not, in fact, appear to be critical in the determination of the optimal operating policy. The dependence was on the variance of the total daily change in cash balance.

Each transaction was of the same nature. With probability 0·5 the cash balance increased or decreased by an amount m. Again m only appeared in critical expressions implicitly in the variance of the total daily change in cash balance.

Miller and Orr specified a policy which constrained the cash balance to lie between two values, the lower one being taken as zero. As soon as the balance reached one of these values it was returned to a single value within the band by means of an asset transfer. The upper bound was denoted by h and the level to which the cash balance was restored after a transfer by z. Note that, because of the Markovian nature of the cash balance process, there must for optimality be a single return point after transfer and not different ones for the two distinct types of transfer. This result was pointed out by Weitzman [30] and Eppen and Fama [5], and it clearly holds if the costs of the two types of transfer differ. The one assumption that must hold, however, is that the costs must be independent of the amounts transferred, i.e. $\gamma(j) = \gamma$. The case of linear dependence was considered in subsequent models, and it was then shown that the single return point policy was no longer optimal.

Within the Miller–Orr policy, optimal values of the h, z parameters are

$$z^* = \left(\frac{3\gamma\sigma^2}{4\nu}\right)^{\frac{1}{3}}, \quad h^* = 3z^*$$

where $\sigma^2 = m^2 t$, the variance of the total daily demand. Note that the return point z^* is not the mid-point of the range, despite the symmetry in the assumed behaviour of the process. Weitzman [30] also showed the way in which the relative values of h^* and z^* differ, depending on the relative costs of the two types of transfer. For example a transfer cost of zero for cash into liquid assets resulted in $h^* = z^*$(cash is kept at the upper limit for as long as possible once it has reached this limit. This was more a case of avoiding the cost of transferring from liquid assets back into cash). In general it would take a transfer cost of assets to cash more than five times that of cash to assets before z^* exceeded $\frac{1}{2}h^*$. The interesting point was that these results were independent of the value of ν and only depended on the ratio of the transfer costs. The optimal average cash balance under the Miller–Orr model was

$$\overline{M}^* = \frac{4}{3}\left(\frac{3\gamma\sigma^2}{4\nu}\right)^{\frac{1}{3}}.$$

Miller and Orr discussed the limitations of their model but made the broad case that, for a relatively simple model, the return was good value. Local information, for example, on the pattern of cash flows (e.g. dividend, tax payments) could be used, but at a cost of developing and analysing a more sophisticated model. For example, they pointed out that if a company knew that it had a substantial payment to make in the near future it may ignore the decision rule which required

Vol. 4
1976

a transfer to short term assets, on the very reasonable grounds that a reverse transfer would shortly have to be made to meet the payment. Conversely, the knowledge that no great outflow of cash was required in the near future may have led management to transfer cash to short term assets when the cash balance was below h^*. Thus the limits were interpreted more as zones or guides around which discretion was exercised.

Some justification for the Miller–Orr model has been described by Homonoff and Mullins [12]. Much of their work was concerned with a practical application of the model where they pointed out that although the assumptions of the model may not be strictly valid, they were sufficiently robust for the results to be of use. This is encouraging, although the potential user might be reassured by something more than an empirical justification. For example, the fact that the original parameters m and t only appeared implicitly in σ^2, should not lead to the conclusion that all that need be considered is the distribution of the daily change in cash balance. The parameters may not be needed, but certainly the analysis depended critically on the assumption of the Bernoulli trials. What would be interesting would be to examine the effects of more general distributions of daily cash balance changes, including changes involving serial correlation. Another criticism of this model is that it assumed stability or stationarity of the stochastic process of cash balances. The important feature of modern cash flow management is that the environment is critically unstable, and any policy adopted must detect and react quickly to changes in the environment.

6. VARIANTS OF THE MILLER–ORR MODEL

The Miller–Orr model assumed that there was a fixed cost γ for cash transfers. Clearly there will be cases where this assumption does not hold, and a number of the simplest variants have been examined. In most cases there was a quite dramatic effect on the form of the optimal policy, departing from the single return point of Miller–Orr.

The simplest modification is to assume a cost directly proportional to the amount transferred, as for example tends to happen with brokerage fees. This case was considered by Eppen and Fama [6] who showed that an optimal policy consisted of two return points U and D, $(U < D)$ to which the cash balance was restored by transfer whenever it strayed from the U, D range. In their previous paper [5] they presented strong experimental evidence that, if the transfer costs have a fixed cost as well as a cost proportional to the amount transferred, the optimal strategy was of the Miller–Orr form of two limits (O, h) but with two return levels, one for each limit.

Derivation of optimum parameters within such a policy could be achieved by linear programming methods (for further experimental work see [21, Chapter V]). The optimality of these results was established by Neave [20].

Girgis [10] introduced a shortage cost (penalty for a negative cash balance) and showed that if the sum of the expected holding and shortage costs for a decision period was a convex function of the cash balance at the beginning of the period, the optimum strategy was of the form described by Eppen and Fama [6], but with the additional feature of changing the balance to a level dependent on the initial balance, even when the initial balance was within the limits. Girgis' proofs were based on a dynamic programming approach. There seems, however, to be a danger now that the complexity of the policies may jeopardise their usefulness. The real message is clear—impose limits on the cash balance, and correct to one or two return levels dependent on how closely the transfer costs may be approximated by fixed or variable amounts. A good estimate of the limits and return levels may then be obtained from the relative transfer costs and from the cash transaction distribution.

In the analysis of models considered, transfer costs (and, where introduced, penalty costs) were included in the objective to be minimised but they did not appear in the actual cash flow. Unlike opportunity costs, these costs do actually mean an outflow of cash from the company, and the assumption that has been made, namely that they were charged against an account distinct from the cash balance itself, seems unrealistic. Porteus and Neave [24] demonstrated how the analysis could be modified so that when, for example, transfer costs were proportional to the amounts transferred, the two limit policy was still optimal. The equivalence of the two formulations and their results was shown further by Porteus [23].

Vol. 4
1976

A further variant arose through a more flexible attitude towards the compensating balance requirement. Hitherto, models have taken this as an absolute limit (usually zero), and this is perhaps an over-simplification. Banks, for example, may operate on an average figure for the cash balance, and penalties are incurred only when the average falls below this specified limit. Moreover, there is a fixed interval (1 month) over which the average is calculated, so that a straight rather than moving average is used. Practices may differ, for example, in the United States and the United Kingdom, in the way the compensating balance concept and corresponding bank charges or services are weighed against each other, but in terms of the cash flow model the effect is the same.

A model of the above type was considered by Hausman and Sanchez-Bell [11]. Transfer costs were assumed linear (fixed plus variable) and different for the two directions of transfer. Costs were based on opportunity costs of the difference (if positive) between the average cash balance and the required compensating balance, and on penalty costs through being below the required balance. The problem was structured in dynamic programming form, optimisation being carried out over the decision points within an averaging interval. Limits (as in [6]) were found for each decision point, and the policy formed as a sequence of such limits. Clearly since the final balance for one interval is the starting point for the next, a rigorous treatment should look at the infinite horizon of intervals,

taking into account this dependence of one interval on its immediate predecessor. Since, however, within each interval there are in practice some twenty opportunities for decisions, any extreme initial cash balance position, caused by a late but necessary correction to the average cash balance, could be rectified quickly without any strong adverse effect on the current average cash balance. In practice the departure from optimality would be small and the computational savings considerable. Hausman and Sanchez–Bell concluded with a comparison between their sequential policy and a static Eppen–Fama strategy. Use was made of data quoted by Homonoff and Mullins [12] and a saving of approx. 18 per cent over the best Eppen–Fama strategy claimed. A similar problem for United States bank cash and asset balance management has been treated by Pye [25].

Frost [9] introduced the concept that the services provided by the bank holding the cash balance produce a yield to the firm dependent on S, the level of services. The variable S in turn depends on the cash balance, either in the form of the average or of the minimum, or perhaps the minimum and the range over which the balance is allowed to vary. It follows from this that any optimisation should include the yield from banking services. Unfortunately, as Hausman and Sanchez-Bell [11] pointed out, Frost carried out his analysis on his modified model using a policy of adjusting to a single level of cash balance, rather than to the two levels of the optimal (Eppen–Fama) policy.

Finally a comprehensive unifying approach has recently been proposed by Elton and Gruber [4]. The basic recursive relationship of their dynamic programming formulation was of the form:

$$f_t(i) = \min_j \left[T(i,j) + L(j) + \alpha \int_{-\infty}^{\infty} f_{t+1}(j-x)p(x)\mathrm{d}x \right]$$

where

$f_t(i)$ = minimum expected cost of managing the cash flow from period t until the horizon T, given that the cash balance at the beginning of period t is i.

j = cash balance after transfer (if any) at the beginning of the period decision point.

$T(i,j)$ = transfer cost from balance of i to balance of j. This is assumed to be two linear functions, one for each direction.

$L(j)$ = expected cost of holding the cash balance = expected opportunity cost + expected penalty cost.

α = discount factor.

$p(x)$ = probability density function of x, the change in cash flow in a period. Observations on x are assumed independently and identically distributed.

The Baumol, Sastry, Patinkin, Miller–Orr, Eppen–Fama, and Girgis models are particular cases of this general model. Elton and Gruber demonstrated the optimality of the form of the policies proposed, which in one or two cases had been matters of conjecture. They did not charge transfer costs against cash balance, but this would not affect their analysis. A further extension to their model covered the case where the assets can be at two levels—essentially short and medium term investments. Modifications were made to the T and L functions, but the general approach was unaltered.

One interesting but more specific model was considered by White and Norman [31]. Chronologically this work is out of sequence in this review as it pre-dates all but the Baumol and Patinkin models, and it can therefore claim the original use of dynamic programming or indeed of any form of sophisticated modelling in this field. White and Norman were concerned with the cash flow problem of an insurance firm. Decisions to transfer cash were made in one direction only, that is from cash to short-term investments. The model allowed for overdrafts and for the realisation of short term loans at fixed times. Dynamic programming was used to derive the optimum policy of a single upper bound to which the cash balance was returned on occasions when the bound was exceeded.

Vol. 4
1976

7. THE THREE ASSET CASH BALANCE PROBLEM

The problem mentioned in the previous section where the assets could be invested at two levels had, in fact, been considered in 1971 by Eppen and Fama [7]. The term 'three asset' here includes as before the cash balance as one asset, the other two being called 'bonds' and 'stocks'. The distinction between bonds and stocks as far as the model was concerned was firstly that bonds produced a lower return than stocks, and secondly that the transfer cost was lower for bonds. Thus, it was argued, it paid an institution to keep a buffer of funds in bonds to absorb an intermediate order of magnitude in cash fluctuations. The main income producing asset of a company would be in stocks, the term covering such assets as stock exchange investments, buildings and actual manufacturing plant of the company. Unfortunately, the authors did not consider the cost of selling stocks at inopportune times, since this would clearly influence decisions. Perhaps this additional factor would render the problem intractable, and the model analysed should again be accepted more for the insight that it provided into the problem. Treatment of a model where the value of the assets was subject to unpredictable variation has been given by Kamin [15]. Kamin showed that for a given (reasonable) utility function, the optimal adjustment strategy consisted of sets of pairs of control limits for the ratio of the amounts of each asset.

A final criticism of this model is that it assumed that transfers were instantaneous. This may be reasonable for bonds, but surely the point about medium

term investments is that there is a significant delay before they can be converted into cash.

8. CASH FLOW AND INVESTMENT DECISIONS

A full discussion of the effects of investment decisions on cash flow is beyond the scope of this review. Rather the intention here has been to look at day-to-day cash flows assuming a stable policy of investment. A recent paper by Franks, Bunton and Broyles [8], although not strictly relevant to the above interpretation of the cash flow problem, did provide a link to what appears to be a useful area for investigation. Franks *et al.* used a decision tree to illustrate what might happen in an investment project, considering such possibilities as strikes, price increases, marketing failures etc. With estimates of the probabilities of these events, probability distributions of cash flows could be built up for the varying decisions which management could make concerning the investment project. The 'transient' as opposed to 'long term average' or 'expected value', nature of this approach has great appeal and particular relevance when the parameters of the model may themselves vary.

9. SUMMARY

It is now nearly a quarter of a century since Baumol's pioneering paper was published. During that time many refinements have been made to the model, but, as happens regrettably with much OR research work, published applications have not always followed advances in the theory. It has become more and more difficult to discern what, amongst the mathematics, is practically feasible. Moreover the assumptions of the models have sometimes been obscured by jargon, and sufficient emphasis was rarely placed on the effects that the assumptions could have on the form of the optimal strategy.

What has been achieved, without doubt, is an insight into the problem and into the form of the policy which should be used. Drawing attention to what could be achieved through a regular control of a company's cash balance and short-term liquid assets brings its rewards, and, if the published policies are followed—perhaps in the sense of evolving good rather than so-called optimal decision criteria—the main purpose of OR work in this area would be fulfilled. In the opinion of this reviewer further refinement would be 'icing on the cake', where the icing is made from a somewhat questionable recipe which may or not enhance an otherwise commendable cake.

It is easy to criticise many of the assumptions made in the models, particularly those concerned with the daily cash flow distribution. Claims are made for the robustness of the resultant policies, and this rather supports arguments of the preceding paragraph. A more relevant criticism is that the models did not

Vol.
197

really provide answers to the main problem faced in the management of cash flows. This is the theme which is developed in the remainder of the review.

10. CASH FLOW MANAGEMENT

An efficient transfer policy between the various forms in which a company may keep its assets clearly is of value, but the real decision faced by the management accountant when he reviews his cash flow position is simply what, if anything, does he do next? Basically he is faced with an unstable environment. He cannot plan on an 'averaging-out' effect of his gains and losses, which would be the consequence of basing his policy on expected values within a fluctuating but in the long term stable economy. If it were not for this inherent instability the problem of cash flow management would merely require a model of a form as simple as a two-way flow version of Baumol. If prices, demands, salaries, wages, interest rates and taxation were all fixed for the foreseeable future, there would be virtually no variability in the cash flow. Payments for raw materials, wages, dividends etc., would be made at regular intervals, and similarly if the output is uninterrupted it would not be unreasonable to assume a steady flow of income. A cash balance, presumably in the form of short term assets, would be built up for the larger less frequent (but still regular) payments such as taxation or dividends. Unfortunately, such an ideal world cannot exist, and models should be built to relate to what is, rather than what ought to be.

What makes the management of cash flow more difficult is the unpredictable occurrence of changes in the environment, such as raw material price increases, wage increases or strikes. Furthermore, it is incorrect to assume that the same rate of inflation on all factors simply produces a scaled-up version of the same financial position. Inflation occurs in discrete jumps, often with a time-lag which works against the manufacturer.

The scope for remedial action is admittedly limited. The obvious remedy for cash flow difficulties is an increase in the price of finished goods although even here there is bound to be some delay, not simply because of statutory restrictions and procedures but also because of the payment lead time by customers. Alternatively the firm can attempt to defer payments to creditors, or exert pressure for more rapid payment by debtors. Kirkman [16] indicates how the importance of trade debtors has changed since 1960, increasing steadily for the first 10 years and then appearing to level off. He argues strongly for a much closer examination of the cost of granting credit. Another more immediate way of improving the cash flow is by running down raw material stocks and reducing the level of production.

Perhaps though the greatest benefit that would accrue from cash flow modelling is that of projecting into the future what is otherwise an instantaneous view

Vol. 4
1976

of the situation. For example, a particularly adverse balance may be the result of the coincidence of several local effects, and the company can look forward, barring crises, to a better immediate future. On the other hand it could be the start of a period where the cash flow is out of balance and some positive action is needed before matters will improve. The importance of being able to distinguish between these two diagnoses is clear. In more general terms HF van den Hoven [29] of Unilever said "Knowing that we will have to live in a world of inflation for a considerable time to come, it is vital that we understand the effect of long-term inflation on the structure of our business and world around us. The most important issue, therefore, is how to deal today with the often violent changes in our environment without damaging the longer term development of the business". Unilever have, in fact, a model based on forecasts of cash movements over a three month period. This, with an operational cash flow statement relating the current stock, debtor and creditor positions to the annual plan and budget, provides management with an effective control over their cash flow. A similar approach has been described by Hubbard [13] where cash flow was monitored using weekly deviations between actual cash flow and planned. The deviations were plotted in standard statistical control chart form, and points outside the three-sigma limits were investigated. The drawback to Hubbard's proposal is that comparisons are being made with budgets made 3 or 6 months ago, which, as Sizer [28] pointed out, are of limited value in a period of rapidly rising costs and falling demands. Nor are these comparisons good indicators of future cash flows.

11. CURRENT AND PROPOSED WORK

The idea of a cash flow model is neither new nor complicated. As a short term budget the concept will be familiar to most company accountants. The aim now should be to develop it more as a management tool, where all the variability and decisions are built in so that a picture of the range of cash flow behaviour can be seen.

One such cash flow model for the oil marketing industry has been reported by Baer [1]. Baer produced a daily cash flow model which approached the facility quoted above, but he used it for the more strategic problems of establishing minimum cash balances and the compensating balance requirement. Again, the attitude was to look at outcomes over a long period of time, in this case a year. The results were derived empirically. More recently Hull and Wheeler [14] and Lawson and Raimond [17] reviewed and proposed models showing cash flows on an annual basis. The Lawson–Raimond paper in particular discussed the "What if?" nature of financial modelling where the outputs in terms of cash-flows of a series of assumed relationships between prices and sales may be tested. Finally, a cash flow model for a vehicle importing company was described by

Churchill and Ward [3] showing the effects of the time lags on such items as freight charges, value added tax, dealer payments etc.

A basic working model has been designed and used by colleagues of the author at Loughborough University. This is a fully deterministic model representing a simple manufacturing process buying raw materials and converting them into a finished product. The inputs to the model are the prices of raw materials and finished products, production rate as a proportion of maximum, fixed costs (administration, expenses etc.), wage rates, taxation, bank charge rates, the distribution of lead time for payment by debtors and the opening conditions. All of the regular inputs can be varied month by month in a multiplicative fashion about base values and the model has been programmed for the computer. The time scale is in units of a month, the output being a statement of the cash flow at the end of each month.

The model demonstrates how cash flows vary from month to month and how these variables can be affected by decisions taken and by external factors. The intention now is to build more sophisticated models using a finer time scale and other extensions such as raw material stock levels, treatment and retaliatory action of creditors, and the relationship between price and demand for the finished product.

Vol. 4
1976

ACKNOWLEDGEMENTS

Grateful acknowledgement is made of the helpful conversations and collaboration of my colleagues, particularly Messrs. RHB Condie, JA Piper, ER Shaw and JM Wilson.

REFERENCES

1. Baer W (1972) A cash flow model. *Managerial Plann.* **20**(5), 24–32.
2. Baumol WJ (1952) The transactions demand for cash: an inventory theoretic approach. *Q. J. Econ.* **66**, 545–556.
3. Churchill M & Ward B (1976) How BMW computerised cash planning. *Accountancy* June, 26–29.
4. Elton EJ & Gruber MJ (1974) On the cash balance problem. *Opl Res. Q.* **25**(4), 553–572.
5. Eppen GD & Fama EF (1968) Solutions for the cash balance and simple dynamic portfolio problems. *J. Bus.* **41**, 94–112.
6. Eppen GD & Fama EF (1969) Cash balance and simple dynamic portfolio problems with proportional costs. *Int. Econ. Rev.* **10**, 119–133.
7. Eppen GD & Fama EF (1971) Three asset cash balance and dynamic portfolio problems. *Mgmt Sci.* **17**(5), 311–319.
8. Franks JR, Bunton CJ & Broyles JE (1974) A decision analysis approach to cash flow management. *Opl Res. Q.* **25**(4), 573–585.
9. Frost PA (1970) Banking services, minimum cash balances and the firm's demand for money. *J. Finance* **25**, 1029–1039.
10. Girgis NM (1968) Optimal cash balance levels. *Mgmt Sci.* **15**(3), 130–140.
11. Hausman WH & Sanchez-Bell A (1975) The stochastic cash balance problem with average compensating balance requirements. *Mgmt Sci.* **21**(8), 849–857.

12. HOMONOFF R & MULLINS DW (1975) *Cash Management*. Lexington.
13. HUBBARD CL (1975) On-going control of cash and income. *Mgmt Accounting* (Nat. Assn. of Accounts) **LVII**(2), 20–24.
14. HULL JC & WHEELER BM (1973) Financial planning: terminal case. *Mgmt Today* December, 37–47.
15. KAMIN JH (1975) Optimal portfolio revision with a proportional transaction cost. *Mgmt Sci* **21**(11), 1263–1271.
16. KIRKMAN PRA (1974) The management of trade debtors. *Accountancy* April, 42–52.
17. LAWSON GH & RAIMOND P (1976) Impact of relative price changes on the one-year budget —Part 1. *Mgmt Accounting* January, 18–20.
18. LOCKYER KG (1973) Cash as an item of stock. *J. Bus. Finance* **5**(1), 44–51.
19. MILLER MH & ORR D (1966) A model for the demand for money by firms. *Q. J. Econ.* **80**, 413–435.
20. NEAVE EH (1970) The stochastic cash balance problem with fixed costs for increases and decreases. *Mgmt Sci.* **16**(7), 472–490.
21. ORR D (1971) *Cash Management and the Demand for Money*. Praeger, New York.
22. PATINKIN D (1965) *Money, Interest and Prices*, 2nd End. Harper & Row, New York.
23. PORTEUS EL (1972) Equivalent formulations of the stochastic cash balance problem. *Mgmt Sci.* **19**(3), 250–253.
24. PORTEUS EL & NEAVE EH (1972) The stochastic cash balance problem with charges levied against the balance. *Mgmt Sci.* **18**(11), 600–602.
25. PYE G (1973) Sequential policies for bank money management. *Mgmt Sci.* **20**(3), 385–395.
26. SASTRY AS R (1970) The effect of credit on transaction demand for cash. *J. Finance.* **25**, 743.
27. SIZER J (1974) The company liquidity crises. *Accountant* **171**, 5215, 744–746.
28. SIZER J (1975) How to control budgets. *Mgmt Today* September, 74–75.
29. VAN DEN HOVEN HF (1975) How Unilever is coping with inflation. Fifth European Management Symposium, Davos.
30. WEITZMAN M. (1968) A model for the demand for money by firms: Comment. *Q. J. Econ.* **82**, 161–164.
31. WHITE DJ & NORMAN JM (1965) Control of cash reserves. *Opl Res. Q.* **16**(3), 309–328.

ADDRESS FOR CORRESPONDENCE: *Professor G Gregory, Department of Management Studies, University of Technology, Loughborough, Leics LE11 3TU, UK*

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

A Decision Support System for Banks

RALPH H SPRAGUE, JR[1]

HUGH J WATSON[2]

(*Received December 1975; in revised form June 1976*)

Just recently, information systems with rather unique characteristics have begun to emerge. These systems usually referred to as Decision Support Systems, feature decision models, a data base and the decision maker as subsystems and are specifically oriented to supporting decision making. This article examines Decision Support Systems and their application to banks.

Vol. 4
1976

INTRODUCTION

MOST medium and large-sized banks have effectively and efficiently integrated computers into their normal banking operations. The evolution of computer usage in banks has paralleled that in other industries. First, the routine data processing tasks, such as the processing of checks, are computerized. Next, summary reports which help managers exercise control over daily operations are provided. And finally, the computer is used to supply information that is designed to support upper managerial decision making. The accomplishment of this latter objective commonly demands the integration of decision models into the organization's information system.

It seems fairly certain that banks, especially larger ones, will increasingly move in the direction of information systems that contain decision models as integral components. However, this development will require a new orientation to the use of decision models. No longer will the models be prepared and used on an ad hoc, non-integrated basis. Instead, they will become part of an integrated system composed of decision models, decision maker, and data base. This article presents a conceptual structure for such a Decision Support System for commercial banks.

[1] Professor of Decision Sciences, University of Hawaii, Honolulu, Hawaii, USA.
[2] Associate Professor of Management and Management Science, University of Georgia, Athens, Ga, USA.

EVOLUTIONARY TRENDS

Banks were early users of computers because of the sheer volume of transactions for which automation was particularly appropriate. However, banks were generally somewhat slower to move to information systems which produced integrated sets of management reports, fast inquiry response from a comprehensive data base, and decision making based on management science models. Recently, however, there have been some significant developments in these areas by banks as well as other types of organizations.

There is a growing recognition that an information system which evolves from transactional data processing, based heavily on structured information flows and data storage, will not be sufficient to support decision making in today's complex organizations. Under the terms Management Decision Systems or Decision Support Systems, several major attempts have been made to develop systems that do support decision making. Morton [11] describes a production control system for an appliance manufacturer oriented around a set of CRT terminals. Boulden and Buffa [4] describe an interactive system of models which draw on data bases for decision support, control, and problem solving. A plywood and a forest products manufacturer is used to illustrate the system. Subsequently, this approach has been applied to a wide variety of organizations [3]. Labelled a General Purpose Operating System (GPOS), it is composed of a financial analysis and planning subsystem, a strategic planning and evaluation subsystem, and an integrated data analysis subsystem. The approach is modular and flexible, utilizing over 20 program components and over 200 mathematical–statistical functions to build and operate a Decision Support System. Over time, this system has become quite popular for long-range planning in the utility industry [7]. Xerox of Canada, Ltd. has been using a system of models implemented in the APL timesharing language which simulate marketing, service, and distribution activities to aid in short-range forecasting and decision making. Recent expansions to this system have resulted in models for operating plans and long-range plans, as well as statistical forecasting models, econometric models, and an extensive series of on-line data bases [14]. Hamilton and Moses [8] describe a Decision Support System based on large scale optimization models and a user-oriented language for modifying and updating them. As a final example of a Decision Support System, Bruce [6] describes a planning and reporting system at Liberty National Bank. The system uses an integrated set of models with a data base which generates forecasts, recommends asset allocation decisions, and graphically portrays past and expected future performance. The system is based on a planning package marketed by Olson Research Associates [20], various aspects of which have been published earlier [15, 16].

Characteristics

None of the systems mentioned above are completely alike. Rather, each is

aimed at specific problems or tasks, usually in a specific industry. Nevertheless, they have much in common with respect to objectives and approach. Generalizing from these early Decision Support Systems, certain characteristics can be identified that seem to indicate a developing trend. These characteristics then point the way for future expansion and refinement. In general, we should expect Decision Support Systems (DSS) to have the following characteristics.

1. The DSS is designed specifically to *support decision making*. Attention to information flows, report structure, and data base design is specifically related to this primary objective.
2. The DSS is interactive to allow the manager or his representative *fast access* to models and data. The interactive capability is not necessarily to provide immediate access to minutes-old data, but, rather, to give access to data and models at a speed which matches the thought processes of the manager.
3. The DSS is *flexible* enough to satisfy the decision making requirements of many types of managers—those in different functional areas, at various managerial levels, and with different management styles.
4. The DSS is an *integrated* set of data and models which allows the models to work together, and thus avoid suboptimization whenever possible.
5. The DSS is *dynamic* enough to keep itself up to date without major or frequent *ad hoc* revisions.
6. The DSS is *sophisticated*, utilizing modern information processing and management science techniques whenever appropriate.

Vol. 4
1976

AN OVERVIEW

The elements of the decision support system are shown in Fig. 1. Notice that there are three major subsystems—decision models, data base, and the decision maker. These major subsystems are marked off by dotted lines and are also labeled. The decision models subsystem consists of the elements that tend to be in the upper right-hand quadrant of the diagram. The data base subsystem is represented by the elements on the left. The decision maker and the means by which he communicates with the system are represented by the elements at the bottom of the diagram.

Decision models

Note that conceptually there are three levels of models. These models are designed to support the primary decision making responsibilities of top, middle, and lower management. Anthony has classified these responsibilities as *strategic planning*, *managerial control*, and *operational control*, respectively [1]. These responsibilities differ considerably and the models that are developed to support these managerial levels must also vary. Decision models for these different

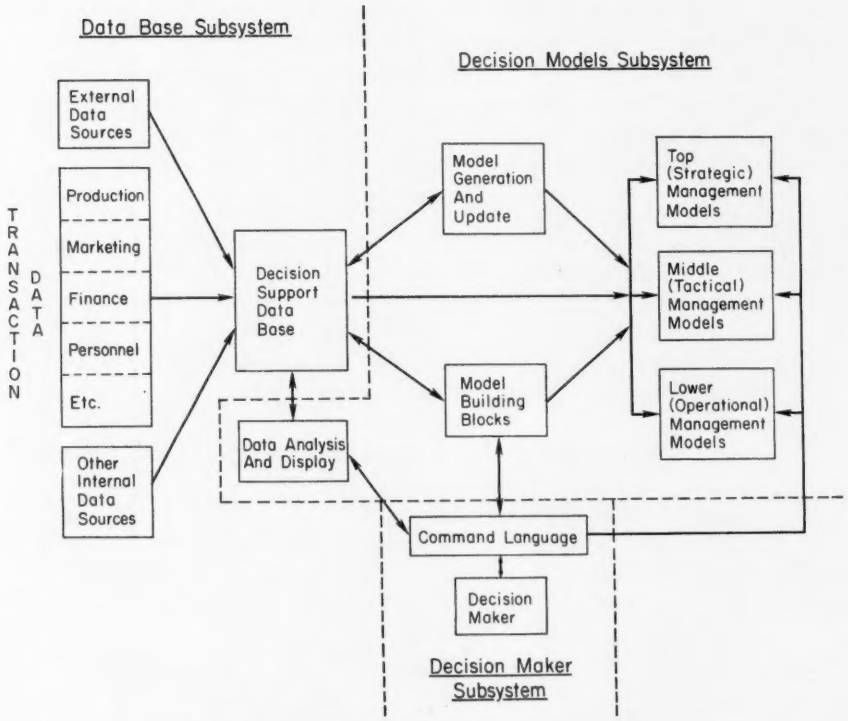


FIG. 1. *Conceptual structure of a decision support system.*

managerial levels are commonly referred to as *strategic*, *tactical*, and *operational models*.[3] These models need communication linkages if they are to be used together to avoid suboptimization.

A major problem with many decision models is that once they are built, they are difficult to change. Consequently, the models gradually lose relevance in the face of the changing decision making environment they were designed to serve. The purpose of the model generation and update is to keep the decision models up-to-date.

Aiding in the building of the decision models are the model building blocks. In general, these are not complete models, but rather, modeling aids, probably in subroutine form. Examples include time series analysis, multiple regression analysis, linear programming, internal rate of return calculations, and so on.

The ability to interrogate the data base, to have the data analyzed in specified ways, and to have the output displayed in a useful format is an important and

[3] These terms have military origins, but have been widely adapted for business use. Generally, 'strategic' refers to major, long-term directional decisions (what to do), while 'tactical' refers to shorter-term resource allocation decisions (how to do it).

desirable feature. The data analysis and display component provides this capability.

Data base

All three levels of models require access to data. The data for the models come from the organization's transaction data base, *and* from other internal and external data sources. An example of internal data would be management's subjective expectation as to future interest rates, while an illustration of external data would be industry loan-to-deposit ratios. The data from these multiple sources form the decision support data base. The dynamic nature of the system and its use requires the data to be continuously updated to reflect recent occurrences.

Decision maker

The decision maker must be able to interact with all of the system's elements. A 'command language' must be available that allows him to direct the actions and steps taken by the system. A wide variety of capabilities should be at the decision maker's disposal. Some capabilities will be required regularly as part of a periodic decision making cycle, while others will be used only occasionally on demand. In general a broad spectrum of mechanisms must be available to the decision maker so that he can use all of the system's capabilities to support his decision making process.[4]

Vol. 4
1976

A DECISION SUPPORT SYSTEM FOR BANKS

Given this introduction to a conceptual structure for a decision support system, let us consider an illustration. In our example, the commercial bank employs a strategic model for the long-range forecasting and evaluating of alternative plans of action. The model is aggregate and macro in focus, and econometric in nature. The major outputs are a 5–10 year forecast of demand deposits and the anticipated demand for loans. Loan demand is categorized into residential mortgage loans, construction loans, installment loans, business loans, etc. Inputs to the model are historical data on these variables from the bank's financial records, and forecasts of economic conditions in the market area in which the bank operates. The long-range planning model also includes management input for major plans and directional strategies the bank wishes to evaluate. For example, the introduction of a new major credit card service aimed at consumers is assessed in light of the model's forecast of retail consumption levels in the bank's marketing area. Figure 2 symbolically shows the flow of data between its sources, the decision support data base, and the strategic model.

[4] For a more detailed discussion of the elements of Fig. 1, see [18, 19].

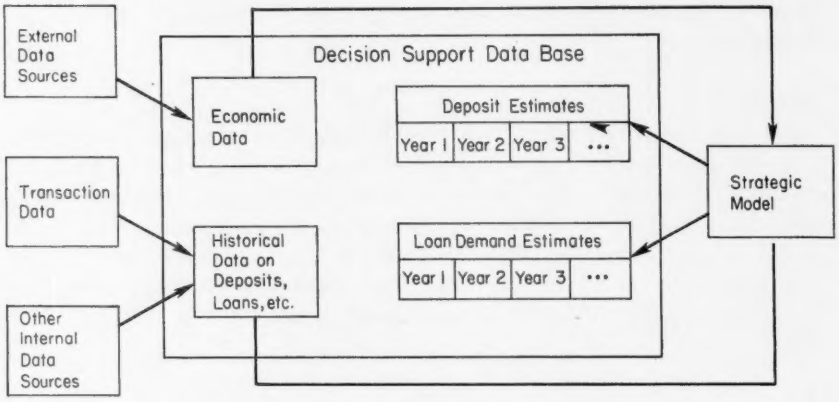


FIG. 2. *Data flows for the strategic model.*

At the tactical level, the bank utilizes a financial planning and control model to forecast the performance of the bank in terms of its balance sheet and income statement for the coming twelve months. This model is important in the month-to-month funds management decisions required of investment and loan officers, and serves as a control device by comparing actual against planned performance. The tactical model's basic variables are those found in the monthly balance sheet and income statement. It maintains a 24-month history data base in financial statement format, and projects 12 months forward using the same format. At the end of each month the new data are inserted in the data base and the oldest month's data are discarded. At that time, the model uses the updated historical data in regenerating the forecast for the coming 12 months. The model draws upon a linear programming subroutine which allocates funds across lending and investment categories based upon the tactical model's estimates of available funds. Recommendations are developed to maximize profitability within constraints on liquidity and capital adequacy set by regulatory agencies and bank management.[5] Figure 3 symbolically represents the data flows for the tactical model.

At the operational level there are many models that the bank can use to support short-range operational decisions. For example, a credit scoring model develops guidelines for the granting of consumer credit loans based on the characteristics and past records of borrowers.[6] The model analyzes past installment loans, both repaid and defaulted, and develops recommended guidelines for granting loans. When new loan applications are coded and submitted to the model, it generates recommended actions based on the guidelines.

[5] For a further description of this tactical model and the linear programming sub-model, see [15].

[6] For an example of this type of model, see [12].

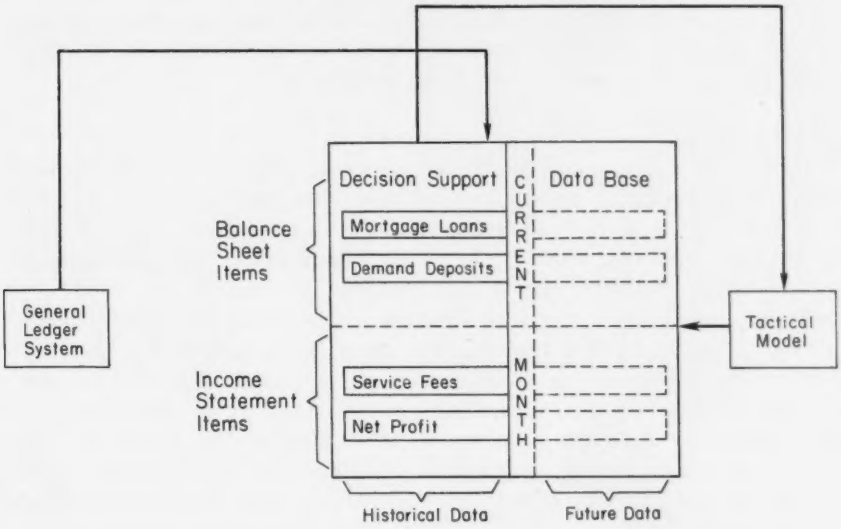


FIG. 3. *Data flows for the tactical model.*

A second example is a bond selection model that analyzes a portfolio and recommends the purchase or sale of bonds.[7] The inputs include the inventory of bonds currently held, a maturity schedule, similar data on other bonds available

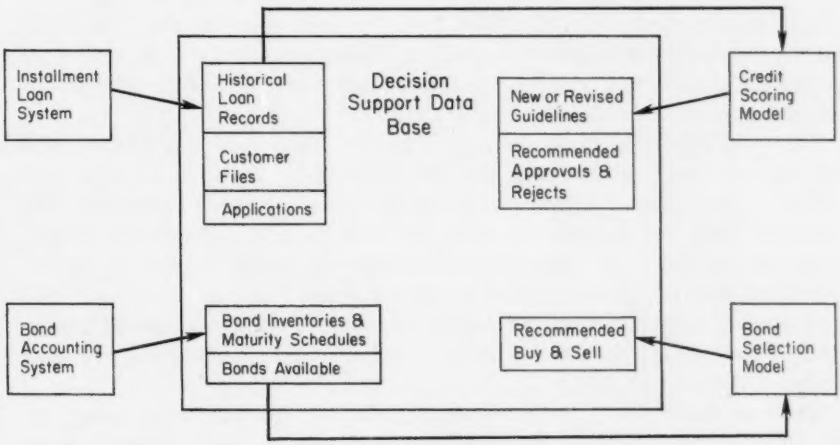


FIG. 4. *Data flows for the operational model.*

[7] For an example of this type of model, see [5].

Vol. 4
1976

for inclusion in the portfolio, and other relevant data. Output is a list of recommended actions that result from the model's analysis. Figure 4 symbolically represents the data flow for the operational models.

Model interfaces

These models need to be implemented in an integrated manner if the overall objectives of the bank are to be pursued effectively. If they are used autonomously or on an *ad hoc* basis, there is significant danger that they might actually work against each other. The result can be good decisions in narrowly defined areas but conflicting or self-defeating decisions in a broader context.

The necessary integration and coordination of model usage is largely achieved through the data base. The importance of integration is evidenced by the need for data flows between the strategic and tactical models. For example, the strategic model's estimate of total deposit levels and loan demand levels for the coming year are necessary inputs to the tactical model. The forecast of total deposits is an annual figure that is converted to an estimate of monthly balances by a combination of time series analysis and seasonal adjustment procedures. The same basic concept holds for the loan demand estimates, which become constraints on the linear programming model for the allocation of available funds among the lending and investment alternatives.

Data may also flow from the tactical level to the strategic level if the linear programming analysis reveals an area of lending or investment that is particularly profitable. Top management can then use the strategic model to assess alternative strategies for pursuing that opportunity more aggressively.

Operational models also interface with the tactical model. In the case of the credit scoring model, the threshold level for loan approval must be based on the availability of funds for this purpose. If the tactical model is suggesting an extension of consumer lending, the credit scoring model must be adjusted to allow additional loans (and vice versa).

To avoid suboptimization, the bond selection model must also interact with the tactical model by accepting the level of available funds as an important input to the analysis. Expansion or contraction of available funds from the tactical model will significantly affect the bond analysis and selection. Moreover, the tactical model may recommend, for reasons of liquidity or capital adequacy, a shift from one type of bond to another. For example, the LP subroutine may suggest a shift from municipal bonds into US government bonds. Data interfaces of this type significantly increase the comprehensiveness and scope of the models' use.

Data must also flow from the operational models into the tactical model. If the threshold level on the credit scoring model results in many potentially good loan customers being refused, it may signal a need to reconsider the maximum installment loan limit imposed by the tactical model. Likewise, the bond selection model might encounter bonds that are highly desirable, even though their

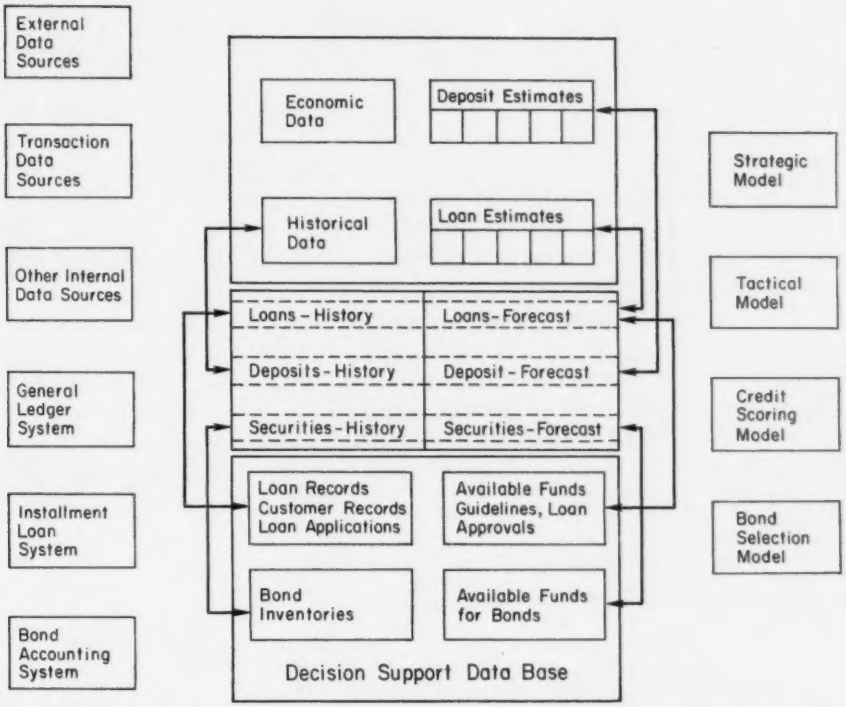


FIG. 5. *Models data and linkages.*

purchase would exceed the recommended limit. This information should be available to the tactical model for reevaluation of the imposed limits.

In short, there must be a spectrum of models to serve the decision maker at the strategic, tactical, and operational levels. These models must work together to avoid suboptimization at any one level or in any one functional area. The mechanism for interfacing them is the systems design philosophy that recognizes and plans for this interaction, and a carefully designed data base to implement it. We have discussed several examples to illustrate the concept. Figure 5 represents the combined system of models and data and shows the linkages between the model levels. For simplicity, only the data linkages which interface the three levels are shown.

Model generation

Models are costly and time consuming to build and later modify. A flexible and powerful mechanism is needed for building and modifying so they do not lose their relevance to the real-world. As an example of a model that is flexible

Vol. 4
1976

(easy to build and modify) and dynamic (responsive to changes in the problem area over time) let us consider in more detail the tactical model discussed previously.[8]

The tactical model can be visualized as a series of equations, one for each entry or line on a financial statement. Each equation develops the value of that line or variable for a given time period based on exogenous variables supplied by some source outside the model, other (previously determined) variables from within the model, and constants and coefficients built into the model.[9] The form of the equation, the variables that enter into each equation, and the coefficients of the equation are determined during the model building stage, usually from an analysis of historical data. Figure 6 shows some segments of this type of model based upon a simplified version of a balance sheet and income statement from a bank. Some of the variables are supplied from outside the model, either as a direct management input (e.g. other expenses) or from another model (e.g. total deposits). Other variables are developed as a function of previously derived variables (e.g. interest expense, total expense) or from a growth rate equation (e.g. salaries and benefits).

Total Deposits (t) = Exogenous (supplied by strategic model)
Demand Deposits (t) = $c_1 \times$ Total Deposits (t)
Time Deposits (t) = Total Deposits (t) − Demand Deposits (t)
.
.
.
Salaries and Benefits (t) = $1.005 \times$ Salaries and Benefits ($t-1$)
Interest Expense (t) = $c_2 \times$ Time Deposits (t)
Other Expense (t) = Exogenous (management input)
Total Expense (t) = Salaries and Benefits (t) + Interest Expense (t) + Other Expense (t)

FIG. 6. *A tactical financial planning and control model.*

This type of model can be easily programmed in a standard procedure-oriented language such as FORTRAN or COBOL. There are some distinct advantages, however, to the use of a 'model definition language' that permits the easy definition and revision of models. Figure 7 shows an example of such a language. Each variable is defined by one of several standard equation forms plus the variables, constants and coefficients that the equation requires. The hierarchical arrangement of the entries (similar to a COBOL data division) shows the determining components of each variable.

The use of a model definition language has several advantages over a typical computer program. First, the logic and structure of the model are revealed. Thus, it serves as a documentation and communication mechanism for people (perhaps

[8] The ideas and concepts illustrated here are also applicable to models of different structures and at the strategic or operational level. For clarity and brevity, we illustrate them here in terms of a simple financial model at the tactical level.

[9] See, for example [2, 10, 21].

01 Total Deposits	EX0	
02 Demand Deposits	E3	c_1
03 Total Deposits		
02 Time Deposits	E1	—
03 Total Deposits		
03 Demand Deposits		
.		
.		
.		
01 Total Expenses	E1	+++
02 Salaries and Benefits	E2	1.005
02 Interest Expense	E3	c_2
03 Time Deposits		
02 Other Expense	EX0	

Equation Forms
E1—Simple combination of terms according to operators shown
E2—Compound growth equation $X_t = cX_{t-1}$
E3—Percentage relationship $X_t = cY_t$
EXO—Exogenous

FIG. 7. *A model programmed in a model definition language.*

even managers) who wish to become acquainted with the model. Second, it permits the model's logic to be changed easily. Equations can be added, the format of the financial statements changed, and the nature of the individual equations modified to reflect changes in the desired output from the model. It is even possible to define the models 'backwards'. A normal sequence is to submit deposits as an exogenous variable and run the model to forecast the net income that will result. With a model definition language, it is possible to structure the model to forecast the level of deposits (or loans) that would be required to generate the desired income level. Finally, a model definition language creates a linkage between models that allows the models to be updated over time. As new data suggests necessary changes in the structure or coefficients in the model, the model logic and coefficients can be stored in a table derived from the definition language, and updated in the data base.[10]

Model update

The initial definition of the model and its equations is often based on data analysis. Subsequent modification and revision of the model require reexamination of the data, particularly to include more recent data that may signal a change in the basic relationships between variables in the model.

There are three modes of periodic update for models. First, the model is rerun with new data inputs, but without any changes in the structure of the equations or the values of the equation constants and coefficients. An example

[10] IBM's Planning Systems Generator reflects a similar philosophy of flexible model definition; see [13].

is the tactical model as it accepts the data for the most recent month, deletes the oldest month from the data base, and regenerates the pro forma financial statements for the coming 12 months. The new forecasts reflect the influence of the new data but there is no change in the model itself.

Second, periodic analysis of historical data can result in a change in the equation constants and coefficients without changing the structure or the variables in the equation. For example, consider the demand deposit equation in Fig. 6. Because of the changing savings pattern of customers, the value of the coefficient c_1 is based on the analysis of the relationship between demand deposits and total deposits in the past. Changing savings patterns may signal a change in c_1 as new data are added to the data base.

Third, modifications in the basic structure of the equations in the model may be required. For example, an equation to forecast the service fees from demand deposits may be developed by performing a step-wise regression analysis. All the variables thought to be useful in forecasting service fees are initially included, and the analysis selects the best variables for predicting service fees. The initial variable set might include demand deposit balances, the number of checks processed, the ratio of special checking accounts to regular checking accounts, and average account balances. The step-wise regression might eliminate all but the first two variables, giving the following equation.

$$SF_t = a + bDD_t + cCA_t,$$

SF: Service Fees during t;
DD: Demand Deposit level during t;
CA: Check Activity level during t.

Several months later, this step-wise regression should be rerun. With an increasing number of special checking accounts, the ratio of special to regular accounts might become a more important variable, leading to a revised equation —which includes this factor as a variable.

The frequency of these three types of revision to the model varies with their importance and extensiveness. The first type of update occurs monthly in conjunction with the normal planning and control cycle. The changes occur automatically unless specifically suppressed. The second type of change, revision, and reestimation of constants and coefficients occurs automatically with notification to the model users on a quarterly or semiannual basis. Major structural changes like the third type occur only annually as a result of major analysis reruns which indicate that structural changes in the model are necessary. These changes are implemented only on specific action by the model builders. The important point is that there should be periodic use of the models and a scheduled review and 'maintenance' program.

Model building blocks

Several times in the preceding discussion there has been reference to the use of standard or packaged statistical/mathematical models. Let us briefly recognize the general function of a library of such standard models and show how they fit into the overall system. The models form a set of building blocks from which more comprehensive models are constructed. The conversion of the single annual deposit figure into monthly deposit figures is a good case in point. A library model using historical data calculates monthly indices to apply to the annual deposit figure in order to develop monthly estimates.

Another example is the use of packaged regression and time series subroutines to perform the periodic data analysis which updates the constants and coefficients in the tactical model. Finally, the tactical model draws on a linear programming subroutine to perform its asset allocations.

Analysis and display

The design philosophy of the DSS stresses the data base as the heart of the system. Models draw data from the data base and generate additional data which in turn are placed in the data base. There are two main advantages to this approach. First, it supports the interaction between models that is so crucial to the system. Data are available to all the models that need it, and the output from one model can be used as input to other models. Second, all data can be formatted and outputted by a common set of routines for display, report generation, and analysis.[11] With a flexible and powerful set of such routines at his disposal, the user can interrogate the data base for answers to specific questions rather than lifting and abstracting the necessary data from preformatted reports. The tactical model, for example, is often used to assess the impact of alternative investment and lending strategies. If the complete set of financial statements were produced after each run, the voluminous data would be extremely difficult to analyze or digest. Instead, the user has an interrogation mechanism that allows him to access only the relevant portions of the new forecast. He can, for example, access only the net income, or return on investment figures, if those are the crucial measures for evaluating the alternative strategies. Thus, the user may access data and compare model output in a way that specifically supports the decisions currently under study without manually filtering and abstracting data from voluminous reports.

The decision maker

Even though the decision models and data base subsystems have been stressed to this point, it should be emphasized that without the decision maker there would be no need for a DSS. The entire DSS is designed to support managerial decision making.

[11] This concept is reinforced by the increasing inclusion of these functions in data base management software.

Vol. 4
1976

The type of DSS discussed here is intended to serve all managerial levels and to cut across the various functional areas. This broad role demands considerable flexibility as to the type of interaction that the system allows.

Consider top management's use of the DSS. In general, top executives in their decision making initially prefer to deal with macro concepts. Consequently, summary or demand reports are preferred over detailed reports, graphical presentations are well received, and ease of system use is imperative. In regards to this last point, the manager may even choose to have a staff analyst serve as an intermediary between himself and the DSS. Hopefully, however, the command language is simple enough to use so that after an initial instruction period an intermediary is not necessary. After the macro concepts are identified and understood by the decision maker, the DSS should permit more detailed analyses in selected areas. For example, the raw data upon which summary reports were prepared should be readily available to the decision maker. By allowing more careful analysis in selected areas, the manager's confidence, and hopefully, the quality of his decisions is enhanced.

As middle, and then lower management's needs are considered in regards to a DSS, difference are noted. For example, more detailed reports are initially required. Graphical presentations are still desirable, but become less important relative to the data underlying the graphs. And the manager himself is more likely to direct the actions of the DSS without the assistance of an intermediary. All of these factors reflect, of course, differences in the decision making responsibilities at the various managerial levels.

CONCLUSION

As has been indicated, Decision Support Systems are now appearing in a wide spectrum of industries and are being used to support many different types of decision making. The fact that they are being developed and used indicates the practicality of DSS. It should be mentioned, however, that most DSS in existence today are not as comprehensive as the one outlined here for commercial banks. The type of system described here reflects what the authors believe that DSS will evolve towards in banks. There still are, however, many problems that have to be ultimately resolved before DSS become widespread in banks and other industries.

That there is sufficient interest and potential value in the area of DSS is witnessed by the January, 1977, Conference on Decision Support Systems being held at the IBM San Jose Research Laboratory. This conference is sponsored by IBM San Jose Research Laboratory, MIT Sloan School of Management, Wharton Business School and SIGBDP (Association for Computing Machines) and is designed to examine existing DSS, develop lists of major problems, and

discuss promising solutions. Through conferences such as this, and the efforts of individual and group research, the future for DSS appears promising.

REFERENCES

1. ANTHONY R (1965) Planning and control systems—A framework for analysis. Division of Research, Graduate School of Business, Harvard University, Cambridge, MA.
2. ANTHONY TF & WATSON HJ (1972) Probabilistic financial planning. *J. Syst. Mgmt* **23**(9), 38–41.
3. BOULDEN JB (1973) A systems approach to corporate modeling. *J. Syst. Mgmt* **24**(6), 14–20.
4. BOULDEN J & BUFFA E (1970) Corporate models: on-line real-time systems. *Harvard Bus. Rev.* **48**(4), 65–83.
5. BRADLEY SP & CRANE DB (1973) Management of commercial bank government security portfolios: an optimization approach under uncertainty. *J. Bank Res.* **4**(1), 18–30.
6. BRUCE JW (1975) Management reporting system: a new marriage between management and financial data through management science. *Interfaces* **6**(1), 54–63.
7. DIETZ RV & SCAVULLO RV (1974) A system to develop utility planning systems. *Public Util.* **93**(10), 59–63.
8. HAMILTON WF & MOSES MA (1974) A computer based corporate planning system. *Mgmt Sci.* **21**(2), 148–159.
9. JACOBS DP, LERNER EM & MOAG JS (1971) Guidelines for designing bank planning models and information systems. *J. Bank Res.* **2**(1), 14–24.
10. KINGSTON PL (1973) Concepts of financial models. *IBM Syst. J.* **12**(2), 113–125.
11. MORTON MS (1971) Management decision systems: computer based support for decision making. Graduate School of Business, Harvard University, Cambridge, MA.
12. ORGLER YE (1971) Evaluation of bank consumer loans with credit scoring models. *J. Bank. Res.* **2**(1), 31–37.
13. Planning systems generator II general information manual. Form No. GH 20-1035, IBM Corporation, Data Processing Division, White Plains, New York.
14. SEABERG C (1973) *On-Line Corporate Models: Tools for Managerial Decision Making at Xerox of Canada, Ltd.* Xerox of Canada, Ltd., Montreal, Canada.
15. SPRAGUE RH (1971) Conceptual description of a financial planning model for commercial banks. *Decision Sci.* **2**(1), 66–80.
16. SPRAGUE RH (1972) Systems support for a financial planning model. *Mgmt Acc.* **53**(6), 29–34.
17. SPRAGUE RH, SASAKI M & SATO M (1974) Computer based management decision systems: conceptual structure and a prototype system. Japan–American Institute of Management Science, Honolulu, Hawaii.
18. SPRAGUE RH & WATSON HJ (1975) MIS concepts part I. *J. Syst. Mgmt* **26**(1), 34–37.
19. SPRAGUE RH & WATSON HJ (1975) MIS concepts part II. *J. Syst. Mgmt* **26**(2), 35–40.
20. *The Financial Planning System.* Olson Research Associates, Inc., College Park, Maryland.
21. WATSON HJ (1972) Financial planning and control. *Mgmt Adviser.* **9**(6), 43–48.

ADDRESS FOR CORRESPONDENCE: *Professor HJ Watson, Department of Management, The University of Georgia, College of Business Administration, Athens, Georgia 30602, USA*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

The Coordination of Short-Run Decision-Making With Long-Range Planning

DENNIS W LOUGHRIDGE[1]

WILLIAM W DAMON[2]

DAVID W PETERSON[3]

(Received October 1975; in revised form June 1976)

Vol. 4
1976

The dichotomy of short-run and long-run decision-making in economic enterprises has gained widespread acceptance, both in practice and in the related literature despite the inherent interdependence of these decisions. The problem addressed in this paper deals with the proper coordination of short- and long-run planning. Structuring the overall decision problem of the firm as a dynamic programming problem, we find that a natural decomposition into long- and short-range planning results. In this decomposition, all decisions which must be made in the first segment of the planning period are allocated to the short-run decision-makers and all subsequent decisions to the long-range planners. Though several formats have been suggested in the literature for linking short- and long-run plans, most notably the use of shadow prices and of targets or quotas, we find that from a conceptual viewpoint these are inadequate mechanisms and that the proper connection involves the dynamic programming return function.

I. THE COORDINATION PROBLEM

THE INHERENT interdependence of the decisions made in a business enterprise has long been recognized. Peter F Drucker [4], in calling for a reorientation of management science in 1959, emphasized this point as follows:

> If there is one fundamental insight underlying all management science, it is that a business enterprise is a system of the highest order . . . .
>
> And one thing characterizes all genuine systems; whether they be mechanical like the control of a missile, biological like a tree, or social like a business enterprise: it is interdependence.

[1] Manager, Division of Business Planning, Allied Chemical Corporation.
[2] Department of Economics and Business Administration, Vanderbilt University, Nashville, Tennessee.
[3] Graduate School of Business Administration, Duke University, Durham, North Carolina.

While this interrelatedness has been accepted, the realization that innate differences exist between what are normally called long-range planning problems and short-run decision problems has also become accepted. As formal recognition of these differences, many business firms have established long-range planning groups which are distinct from groups responsible for the short-run operation of the firm. This contradiction is emphasized in the following from Anthony [1]:

> Many actions that are taken for long-run strategic reasons make current profits smaller than they would otherwise be. . . . The problem of striking the right balance between strategic and operating considerations is one of the central problems in the whole management process.

A variety of budgetary mechanisms have found widespread use as instruments of coordination between long-run and short-run decision-making. These budgetary mechanisms are goal directed since the budget establishes goals on the level of achievement of the various decision-makers. These goals may be based on physical factors, such as inventory levels, or on monetary or accounting related factors, such as profits or return on assets. Goals could be established for each entry in the firm's balance sheet in which case the budgetary process would be aimed at directing the firm to the "best possible end-of-the-period balance sheet" (see [5]).

An alternate mechanism for coordinating the decision-making activities of interrelated groups is transfer pricing. While this mechanism is normally thought of as a means of coordinating divisions or profit centers within a firm it may also be applicable to the coordination of short-run and long-run planning. In this use, transfer prices would be established on transfers between time periods.

The purpose of this paper is to evaluate these and other mechanisms which may be appropriate for coordinating long-run and short-run decision-making.

II. FRAMEWORK FOR ANALYSIS

If it were possible to collect all relevant information in a single location, and to develop and solve a model which could simultaneously determine the optimal values for all decision variables then it would not be necessary to decentralize the firm. Although such centralized operations are not generally feasible, an ideal centralized decision model is valuable as a standard against which to measure the effectiveness of a decentralized system for making the firm's decisions. The particular decentralization of interest here is that in which long-range decisions are allocated to one group, and short-range decisions to a second. It is the effectiveness of this type of decentralization we wish to examine.

The following taxonomy is useful in describing the problem which would be faced by the single decision-maker in a centralized model, and also the rela-

tionships among his problem and those of the long- and short-run decision-makers in a decentralized organization.

1. Decision variables: those variables over which the firm has direct control, examples being the production quantity for the coming time interval, immediate outlays for advertising, and borrowing or lending decisions.
2. State variables: those variables over which the firm has indirect control through its decision variables such as the level of inventory (which is influenced by the production quantity decision), production capacity, debt and equity levels, and market share. The state variables are so chosen as to constitute a description of all characteristics of the firm which are relevant to its future performance. This complete description at any instant of time is referred to as the state of the firm at that time.
3. State transition equations: a statement of the influence of the current state and decision variables on the values of the state variables in the immediate future. Thus, the inventory balance equation (New Inventory = Old Inventory + Production Quantity − Sales) is a state transition equation.
4. Constraints: limitations imposed on acceptable values of the state and decision variables. For example, inventory must not exceed warehouse capacity.
5. Objective function: the scale on which the firm's performance is measured.

Vol. 4
1976

To apply these concepts to the problem faced by the decision-maker of an ideally centralized firm, we assume that the firm's planning horizon can be divided into a series of N discrete time periods, say years.

In period t ,the decision-maker must choose numerical values for the decision variables $D(t)$, based on his knowledge of the system state $S(t)$, and on his assessment of the impact his decision will have on subsequent states of the system. In consequence of the system's state at time t and the particular decisions $D(t)$, a payoff $P(t)$ for that period is realized. The index of overall performance may be the sum or discounted sum of the payoffs $P(1), \ldots, P(N)$.

Decisions implemented in any period thus have a two-fold effect on the firm's performance. The first is a direct impact on current performance, the second a dynamic impact on future performance. This latter occurs through the effect of immediate decisions on the future state of the firm, which in turn affects future performance. For example, consider a production manager who must choose between current production and general maintenance. The decision reached affects both current and future performance.

There is no issue of coordinating short-range and long-range plans for the ideal decision-maker. His problem is the determination of the vectors of decision variables, $D(1)$ through $D(N)$, which optimize the corporate objective function over the entire planning horizon. All tradeoffs between long-range and short-range performance are made by the ideal decision-maker in accordance with the corporate objective function. The usefulness of this conceptualization lies not in its direct applicability, but in its use as a basis with which to compare other decision-making structures.

III. THE DYNAMIC PROGRAMMING APPROACH

Numerous mathematical techniques are available which are applicable to the type of problem faced by the ideal decision-maker. Of these, dynamic programming offers the advantage of providing a natural decomposition of the decision problem into short-run and long-run segments. The solution technique is based directly on Richard Bellman's Principle of Optimality. In Bellman's words [2]:

> "An optimal policy has the property that whatever the initial state and decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

The principle states that it is not necessary to consider the firm's past decisions in order to determine the optimal decisions for a particular period. This simplification is made possible by the use of state variables which summarize all the information about the history of the firm which is relevant to its current decision problem. To illustrate this concept, consider the single period version of the ideal decision-maker's problem in Fig. 1. All the information about the history of the firm is specified by $S(1)$. The state of the firm at the end of this period, $S(2)$, will provide an updated statement of the status of the firm for use in the following period's decision problem.

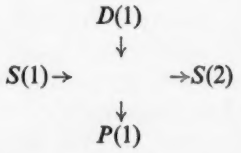


FIG. 1. *Single period version of ideal decision-maker's problem.*

Since the state of the firm at the beginning of the period being planned for may not be known at the time of the formulation of the plan, dynamic programming determines decision rules which specify the optimal decisions to be followed for each possible initial state. Another concept of dynamic programming which will play a central role in addressing the decomposition and coordination problems is that of the optimal return function. If an optimal decision policy is followed from time $t = T$ onward, the performance of the firm thenceforth is a function (the optimal return function) only of the state of the firm at time T. In the single period example, the optimal return function specifies the best possible performance the firm can achieve for each possible initial state, $S(1)$.

An example of the applicability of these concepts is a control system designed to bring the firm from some initial statement of the balance sheet accounts to some desired end-of-the-period balance sheet at minimal cost ([5], p. 208). The entire balance sheet is the vector of state variables and the planning problem consists of jointly determining the desired terminal balance sheet and those

transactions (resulting from decisions) which will realize the terminal balance sheet at minimal cost.

The Principle of Optimality implies that this problem can be solved without consideration of the decisions employed to reach the state described by the initial balance sheet. Furthermore, the performance of the firm, given that the optimal decision rules are followed, is dependent only upon its initial balance sheet if the terminal balance sheet reflects the impact of each of its components on the future performance of the firm. Thus, the determination of the desired terminal balance sheet is a critical part of this planning process.

This example demonstrates how the taxonomy of dynamic programming can be imposed upon one type of planning system. Before discussing the merits of this or other planning systems, it is useful to demonstrate the solution technique of dynamic programming.

Suppose that a large United States firm owns a coffee plantation in a South American country. The political situation in the country is such that the managers of the firm believe that the plantation will be expropriated without compensation in 3 years. A market does exist for the four identical sections of land included in the plantation. However, because of general uncertainty over the future of private enterprise in the country, the expected value of the land decreases rapidly over the coming years. It is also believed that any attempt to unload the land at a rapid rate would increase skepticism and thereby further reduce the selling price. Thus, the marginal value to the firm of selling additional units of land in any one time period decreases. Table 1 provides management's best estimate of the selling price for the simultaneous sale of N sections of the land at the present time and at the beginning of the subsequent years. The corporate planning group must determine the optimal policy for disposing of the land, given that continued use of the land contributes \$100,000 of profit per section of land per year.

Vol. 4
1976

TABLE 1. ESTIMATED LAND PRICES

Sections of land sold	Total selling price (in \$100,000 units) year 1	2	3
1	8	6	4
2	14	10·5	7
3	18	13·5	9
4	20	15	10

This problem may be summarized as follows:

1. Decision variable: number of sections of land to be sold at the beginning of each year.

2. State variable: number of sections of land unsold at the beginning of each year (before sales for that year).
3. State transition equation: function which specifies that the number of sections of land owned at the beginning of a year must equal the number of sections owned at the beginning of the previous year minus the number of sections sold in that period.
4. Constraint: relationship which limits the number of sections of land sold in any period to be less than or equal to the number available to be sold.
3. Objective function: profits over the 3-year planning horizon.

Since all land is expected to be expropriated at the end of the 3-year period, the land still owned by the firm at that time has no value. The decision problem for the last year thus reduces to maximizing the profits for that period. However, the state of the firm at the beginning of the last period is not known so a decision rule is required which will specify the number of sections of land to sell for each possible initial state. Since the firm began with four sections of land, it would be possible to own 0, 1, 2, 3 or 4 sections at the beginning of the last period. For any of these states the optimal decision rule and the corresponding optimal profit in the third year is specified in Table 2.

TABLE 2. OPTIMAL DECISIONS AND PROFITS FOR YEAR 3

Sections owned at the beginning of year 3	Optimal number of sections to be sold	Optimal Profits (in $100,000 units)
0	0	0
1	1	4
2	2	7
3	3	9
4	3, 4	10

The specification of more than one optimal decision indicates that alternate optimal solutions exist. In this case, if four sections of land are owned at the beginning of the last year, and the firm sells all four, the profit resulting from the sale will be $1 million. However, if four sections of land are owned and the firm sells three and employs the fourth, the profit will be $900,000 from the sale plus $100,000 from production for a total of $1 million.

Table 2 specifies both the optimal return function and the optimal decision rule. The optimal return function specifies the maximum profit which will be realized in this and all future periods for any possible state, given that the optimal decision rule is followed.

The next period to be considered is year two, the next to the last year. In this period, the decision-maker must consider not only the immediate profits which result from his decision but also the state in which his decision leaves the firm. The relevance of the future state of the system can now be stated directly: it

determines, via the optimal return function, the profits which will be earned in the last period.

In order to insure that the period two decision-maker considers both the long– and short-run implications of his decision, he should be charged with maximizing total profits over the last two periods. The decision problem has, however, been simplified since it is not necessary to consider explicitly the decision which will be made in the final period. All the information required about the impact of period two decisions on final period profit is provided by the optimal return function.

For example, suppose that at the beginning of period two, two sections of land remain unsold. If the decision-maker were motivated to maximize only current profit, he would sell both sections, producing a profit of $1,050,000 in the current period. Since no land would be available for the last period, the total profit for the last two periods would be $1,050,000.

Consider, however, the results which occur if the period two decision-maker had full knowledge of the optimal return function for the last period and were charged with maximizing total profit for the last two periods. Again assuming that two sections of land are available at the beginning of period two, the optimal decision would be to sell one section this period. This would produce a profit of $11 million over the last two periods as shown in Table 3.

Vol. 4
1976

TABLE 3. CALCULATION OF PROFIT FOR YEARS TWO AND THREE (MONETARY AMOUNTS IN $100,000 UNITS)

Period two	
Profit from selling one section of land	$6
Profit from employing one section of land	1
Period three	
Maximum profit given that one section of land is available	4
Total profit	$11

The period two optimal return function and decision rules are provided in Table 4.

TABLE 4. OPTIMAL DECISIONS AND PROFITS FOR YEAR 2

Sections owned at the beginning of year 2	Optimal number of sections to be sold	Optimal profits in $100,000 units
0	0	0
1	1	6
2	1	11
3	2	15·5
4	2	19·5

The number of sections of land to be sold in the first period must now be determined in such a way that total profit over the 3-year planning period will be maximized. The decision problem in the first period is somewhat simplified since the initial state is known; i.e. four sections of land are owned.

The total profits for all possible period one decisions are presented in Table 5 where it is seen that the sale of two sections in period one will lead to maximum profit.

TABLE 5. MAXIMUM TOTAL PROFITS

Sections of land sold in period 1	Maximum total profits in ($100,000 units)
0	19·5
1	24·5
2	27
3	25
4	20

The purpose of this example is to illustrate the dynamic programming approach to solving sequential decision problems. The significance of the approach is that the sequential decision problem can be divided into stages such that only one decision (or one set of decisions) is required at each stage. As was demonstrated, the only information which must be passed between stages in order to insure that an optimal solution is reached is the optimal return function.

IV. IMPLICATIONS FOR COORDINATION

Consider once again the ideal decision-maker's problem as a sequential decision problem. By applying the principles of dynamic programming, it is possible to dichotomize the planning problem as shown in Fig. 2.

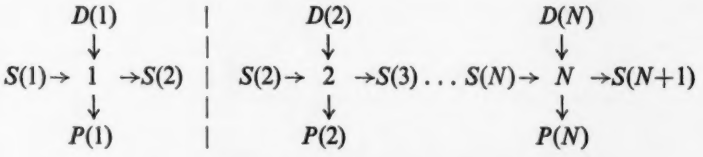


FIG. 2. *The dichotomized planning problem.*

In this dichotomy, all decisions to be made in the first period are allocated to the short-run decision-makers and all decisions to be made in subsequent periods are allocated to the long-run decision-makers. As in all cases when interrelated decisions are allocated to different decision-making groups, some mechanism

must be established to coordinate the activities of each group. The goal of the coordination mechanism is to ensure that the performance achieved by the decomposed structure is as close as possible to that achieved by the ideal decision-maker.

One possible means of coordinating the short-run and long-range planners is to assign a value to each of the possible terminal states for the short-run problem. In order to ensure that optimal short-run decisions are made, the values assigned must reflect the best possible future performance the firm can achieve as a result of being left in a particular state. This value is precisely that generated by the optimal return function for the long-range planning problem.

Dynamic programming thus leads to a straightforward solution to the co-ordination problem. The role of the long-range planners is to determine the optimal return function for the long-run planning period, beginning with the unknown initial state $S(2)$. Once this optimal return function has been determined, it is used to inform the short-run decision-makers of the long run impact of every possible current decision.

An important advantage offered by this means of co-ordination is that short-run decision-makers do not have to consider explicitly future decisions. The optimal return function provides the short-run decision-maker with sufficient summary information on the long-run impact of short-run decisions to permit him to make the tradeoffs between the two planning problems. The following quotation from Drucker [3] concerning long-range planning is particularly relevant to this type of coordination mechanism:

> It (long-range planning) does not deal with future decisions. It deals with the futurity of present decisions.

Two alternate mechanisms for coordinating the long-run and short-run planning problems are suggested by dynamic programming as well as current business practices. The first of these requires the long-range planners to specify a desired terminal state for the short-run decision problem, $S(2)$.

The specification of a required terminal state is conceptually similar to budgetary controls used in many companies (see [5]). In addition to the specification of a desired terminal balance sheet, targets are often set on a wide variety of other factors such as inventory levels, cash position, production capacities, and work force levels. The difficulty with this approach is that there is no way to determine what these targets should be without intimate knowledge of both the short-run and long-run planning problems; but this requires a centralized organization.

A further problem is that there may be no feasible solution to the short-run problem which generates the required terminal state. This is apparently the same problem addressed by Schaffer [6] in a more general context:

> In practice, however, the actions required to carry out the long-range plans may not be consistent with the actions required to meet immediate operating goals.

Vol. 4
1976

The second alternate approach to the coordination problem is based on the familiar concept of transfer prices. The short-run decision-maker is given a credit which measures the value of the state in which the firm is left at the end of the planning period. This credit is equal to the marginal value, or transfer price, associated with each state variable times the numerical value of that state variable. To illustrate this approach, the optimal return function for the long-run plan and the marginal values of each section of land from the coffee plantation example are presented in Table 6. The immediate problem encountered is which of these marginal values to use for the transfer price. In general, there is no way to solve this problem without knowledge of the optimal number of sections of land to have available at this time. Consequently, the optimal transfer price is dependent on the optimal number of sections of land available, which is unknown.

TABLE 6. CALCULATION OF MARGINAL VALUES

Sections owned at the beginning of year 2	Maximum profits (in $100,000 units)	Marginal values (in $100,000 units)
0	0	
1	6	6
2	11	5
3	15·5	4·5
4	19·5	4

Consider the first period decision problem in the above example and assume that a transfer price of $600,000 is specified. The profit associated with each possible decision is presented in Table 7. The short-run decision-maker would thus be motivated to sell only one section of land rather than two, which is optimal. Total profits for the entire planning period, from Table 5, would be reduced from $2,700,000 to $2,450,000.

TABLE 7. SHORT-RUN DECISION PROBLEM WITH TRANSFER PRICES (MONEY AMOUNTS IN $100,000 UNITS)

Sections of land to be sold	Profit from production	Profit from land sales	Profit from transfers to subsequent periods	Total profits
0	4	0	24	28
1	3	8	18	29
2	2	14	12	28
3	1	18	6	25
4	0	20	0	20

The only case in which it is possible to determine appropriate transfer prices without knowledge of the optimal terminal state is when the optimal return function is linear, so that the marginal values are constant. Otherwise, the optimal terminal state must be known in order to determine the optimal transfer price.

V. SUMMARY AND CONCLUSIONS

This paper deals with the problem of decomposing an integrated planning model, the ideal decision-maker's problem, into short-run decision-making and long-range planning and of coordinating the resulting segments. The objective is to identify a decomposition and a corresponding coordination mechanism which insures that the performance achieved by the decomposed decision structure is equivalent to that achieved by the ideal decision-maker. This goal is realized primarily by the application of dynamic programming concepts.

The resulting decomposition allocates all decisions which must be made in the first segment of the planning period to the short-run decision-makers and all subsequent decisions to the long-range planners. Three distinct coordination mechanisms were investigated which involved the specification of the following summary information concerning the long-range plans to the short-run decision-makers.

1. A required terminal state for the short-run decision problem.
2. Transfer prices on the terminal state variables for the short-run decision problem.
3. The optimal return function for the firm's long-run planning problem.

The effectiveness of each of these coordination mechanisms is as follows:

1. Specification of a required terminal state (goal) for the short-run problem is sufficient to achieve optimal coordination only under extremely restrictive conditions. These conditions imply that the long-run planners have complete knowledge of the details of the short-run planning problem.
2. Only in the special case of a linear optimal return function is it possible to determine optimal transfer prices for the short-run state variables without also knowing the optimal terminal state.
3. Complete specification of the optimal return function associated with the firm's long-run planning period is sufficient to insure that short-run decisions will be optimal with respect to the firm's overall objective.

These conclusions demonstrate that the appropriate role of long-range planning is the estimation of the firm's optimal return function and the appropriate objective for the short-run decision-maker is the maximization of the performance of the firm over the entire planning horizon given the optimal return function.

These conclusions also point out that the typical techniques for coordinating

Vol. 4
1976

short-run and long-run planning (i.e. the specification of targets values or budgets, and the specification of intertemporal transfer prices) may lead to suboptimal performance for the firm.

REFERENCES

1. ANTHONY RN (1970) Framework for analysis. In *Information for Decision Making* (Ed. RAPPAPORT A), pp. 99–108. Prentice Hall, Englewood Cliffs, NJ.
2. BELLMAN RE & DREYFUS SE (1962) *Applied Dynamic Programming*. Princeton University Press.
3. DRUCKER PF (1970) Long range planning: challenge to management science. In *Information for Decision Making*. (Ed. RAPPAPORT A), pp. 109–118. Prentice Hall, Englewood Cliffs, NJ.
4. DRUCKER PF (1959) Potentials of management science. *Harv. Bus. Rev.* **37**, 25.
5. IJIRI I, LEVY FK & LYON RC (1970) A linear programming model for budgeting and financial planning. In *Information for Decision Making*. (Ed. RAPPAPORT A), pp. 208–219. Prentice Hall, Englewood Cliffs, NJ.
6. SCHAFFER RH (1967) Putting action into planning. *Harv. Bus. Rev.* **45**, 158.

ADDRESS FOR CORRESPONDENCE: *Professor W. Damon, Department of Economics and Business Administration, Vanderbilt University, Nashville, Tennessee 37235, USA.*

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

A Quantitative Model for Manpower Decision Making[1]

HARVEY KAHALAS

DAVID A GRAY

(*Received March 1976; in revised form June 1976*)

Today's society is continually influenced by various changes, requiring managers to plan for manpower requirements. An approach that can be used by firms to solve one of the most pressing of today's employment problems is examined. In effect, the problem becomes one of determining an optimal work force composition based on both sociological and skill components. The problem is viewed as one of establishing rank ordered priorities among multiple conflicting manpower objectives. This study discusses some of the internal and external constraints faced by the firm and suggests a particular technique, goal programming, to facilitate the manpower decision-making process. The technique is examined under two different circumstances to provide some indication of the model's flexibility.

Vol. 4
1976

ALTHOUGH authors differ as to the optimal number of stages in manpower planning, they perceive the same activities being necessary for effective manpower planning. These may be viewed as inventory, forecasting, analyzing organizational effects, programming, and implementation [3, 13, 15].

Implicit in each manpower planning stage is the need for information. The firm must have information concerning both its internal manpower supply and the external manpower sources from which the firm plans to recruit. The amount and type of information collected will depend substantially on the size of the firm, its sophistication, the scope of its manpower planning program, and the executive in charge. These factors will eventually determine the effectiveness of the firm's manpower planning. Herein lies the problem with manpower planning currently being practiced by many firms—many manpower planners are still operating by using *ad hoc* and armchair planning approaches [11].

The reason for modifying this existing approach stems from the rapidity of social change [2, 14]. Within the past decade, there have been sweeping changes in our society with regard to governmental regulations concerning employee

[1] Thanks should be given to Mr. William Hawkins, a graduate student at Virginia Polytechnic Institute and State University, for his help in the computer analysis.

selection which have had a tremendous impact on manpower planning [4]. Additionally, many other factors such as family planning and educational development have created needed modification in our organizations. If the past is any indication of the future, change will continue to occur at an increasing level. As the magnitude of change increases, so too will the instability of the environment in which the firm must operate. Coupled with this increase in instability will come increased uncertainty about the external conditions, and the firm will move even further into a turbulent environment [6, 12].

Change and the uncertainty facing the firm require that manpower planners use more sophisticated techniques and better manpower information [9]. Burack [3, p. 95] suggests that instability and uncertainty ". . . resulting from . . . changes which import unstable conditions to job structures and complicate the (successful) meeting of occupational requirements" are the forces which ". . . argue strongly for the organization of manpower procedures and compel serious consideration of a viable manpower information system . . ." [3, p. 96]. Accordingly, manpower planners must attempt to anticipate and incorporate into their manpower plans environmental changes which they think will affect the firm's manpower planning. The changes which each firm attempts to incorporate will depend on the environment of the firm. A workable manpower planning model must show how the firm can incorporate anticipated environmental changes [12].

The environmental changes which can affect manpower planning are many and varied. Three of the most important factors are technological, demographic, and governmentally induced change [13, pp. 2–5]. By taking these factors into account, it becomes readily apparent that manpower planning is becoming increasingly more difficult and that firms must acquire more and better information through the use of more sophisticated techniques for manpower planning. Specifically, the firm must attempt to incorporate current and expected environmental changes in its manpower decision-making model.

Therefore, the ubiquitousness of societal change affecting human resources has increased the need for the development of methodologies for facilitating manpower decision making. This suggests that a major concern for management might well be the need for methods and models for manpower planning that are sensitive to these various changes. Accordingly, the objectives of this paper are (1) to develop a manpower planning and decision-making model utilizing an application of goal programming and (2) to illustrate the model through generating a set of informational (decision) outputs from hypothetical data.

GOAL PROGRAMMING

Goal programming is a technique in which externalities and social desires can be joined with organizational constraints in the establishment of an ordinal

hierarchy of goals which the decision maker desires to achieve. The basis of this approach is subjective in nature, but it does allow for the ranking and quantification of the manager's values in his organizational role tempered by the need to consider environmental pressures. Some areas where goal programming has been applied as an aid in decision making are production, finance, marketing, corporate planning, academic planning, governmental analysis, and medical care planning [8, pp. 191–350].

The methodology of goal programming treats multiple goals in the same objective function by minimizing the deviations from each specified goal. All constraints are treated as goals by the model with associated priorities so that the values assumed by the deviational variables of a constraint determine the extent of goal achievement in the final solution. Constraints in the model are dichotomized into system and nonsystem factors. The system constraints are those necessary values which management requires to be initially satisfied by the model. The remaining or nonsystem constraints are then treated as goals, and the model attempts solution in a hierarchical order.

The deviational variables of the goals are represented in the constraints and priority equations as d^+ and d^-. If management desires to attempt to minimize the positive deviation, d^+ is the symbol used, while d^- would be used to attempt to minimize the negative deviation. Other, more complex variations in the deviational variables of the objective function also may be employed [8, pp. 40–41].

Vol. 4
1976

Furthermore, goal programming operates on an optimization basis. The model's basic objective is to search for an optimal result which allows for the maximization of goal attainment. Its solution is optimal only given the specific model's objective function and constraints. Therefore, the formulation of these factors must be done carefully and critically to assure the best result for decision making [1, pp. 115–130; 10, pp. 247–264].

A model will be developed in two successive stages. The first stage will be a basic manpower decision-making model, including the development of the governing constraint equations and the objective function which defines the hypothetical company's goals concerning its manpower requirements. Additional constraining equations which are imposed by the company's environment will be included. The second stage will include final constraints in the form of a manpower training program and its budget. Again, the objective function will be constant so that any effect the training budget has on the company's manpower requirements or its original objectives can clearly be identified.

A few basic limitations of goal programming as a quantitative technique have been previously identified [8, pp. 32–34]. The first of these limitations concerns the concept of proportionality. Because of the fact that goal programming is an extension of linear programming, the linearity factors between the objective function, goal relationships, and constraints must be observed. Since the proportional factor does not ensure linearity, the objective function and the

constraints must be additive. These two factors, proportionality and additivity, will basically achieve the goal of linearity. Nevertheless, these do not present themselves as a particularly limiting factor for our model because of its inherent linearity.

A goal programming model can be used to deal with problems concerning the external source of labor as well as the internal sources. The external source allows for inflows into the existing organizational structure while the internal sources are included in the structure and are specifically identified in terms of mobility, via the incorporation of a transition matrix which identifies flows of workers between divisions at some given time period. Therefore, the model attempts to implicitly deal with such concepts as skill transference. Even with this matrix, certain factors such as changes in type of manpower used and various levels of cost behavior cannot be incorporated in the model to gain a range of results. Nevertheless, to allow for some indication of this flexibility the model can be updated by modification in constraints or priorities.

Another inherent limitation is the absence of integer goal programming algorithms. Although the optimum solution may yield fractional results, for the purposes of this model the results can be rounded to the nearest integer value.

The final limitation concerns the deterministic mode of the coefficients in the model. This seems to convey the impression of a static decision environment, but that is not necessarily correct. The use of goal programming in its present stage of development does not allow for probabilistic factors to be incorporated into the model as coefficients. Nevertheless, because of the ability to modify priority ranking structures and/or to restructure the objective function, the model does provide some flexibility to the decision maker. Because of the need for future research in parametric goal programming, some limitations may exist in the model. Nevertheless, the technique still may be useful as a decision-making device [7; 8, pp. 175–185].

A PROPOSED MANPOWER DECISION-MAKING MODEL

The company for which the model is hypothetically designed is a manufacturing firm with two divisions, A and B. Each division requires the employment of skilled workers to operate its machinery. The machinery in division B is more sophisticated than that of division A and requires more skill to operate it properly. Because of the increased skill level in division B, the average salary of the employees in division B is higher than the average salary of those in division A

There are two sources of manpower in this model, source 1 and source 2. Source 1 represents the male source of manpower, and source 2 is the female manpower source. Each source, in turn, can be either external or internal. An external source refers to the company's source of manpower in its environment

or, in other words, the manpower availability in the recruiting area of the company. This is a function of such things as the company's size, its location, the nature of its business, and labor market conditions. An internal source refers to the manpower within the company. This source exists because employees are capable of an allowed movement between the two divisions of the company. Finally, the model will allow for distinguishing sex within the categories of Caucasians and minorities.

Even with these category breakdowns, the gross categories may not be meaningful for most companies. For example, while unskilled labor may have a certain homogeneity, it would seem that the skilled category may contain numerous types of skills. Therefore, it may not be valid to lump manpower into the gross categories of skilled *vs.* unskilled. Nevertheless, for a hypothetical example it does simplify the technique while still allowing enough difference to show its applicability.

Constraint equations

The constraint equations are quite numerous and are developed based upon the variables to be incorporated into the model. In the model, all internal sources of manpower are considered skilled personnel. The first set of constraint equations express the interrelations among the variables as a result of their definitions. These are expressed algebraically as follows:

$$\begin{aligned} x_1 - x_2 - x_3 - x_4 - x_5 &= 0; \\ x_6 - x_7 - x_8 - x_9 - x_{10} &= 0; \\ x_{11} - x_{12} - x_{13} - x_{14} - x_{15} &= 0; \\ x_{16} - x_{17} - x_{18} - x_{19} - x_{20} &= 0. \end{aligned}$$

The next set of constraint equations arises from restrictions on the manpower available in the recruiting area of the company. The company is well acquainted with the breakdown of the work force in its area and can, therefore, accurately estimate the manpower available (MPA) in the various segments of the work force. The constraints have the form, e.g. the number of skilled minorities employed from external source 1 for division A plus the number of skilled minorities employed from external source 1 for division B must be less than or equal to the estimated and available total number of skilled minorities from external source 1, i.e.

$$x_2 + x_7 \leq \text{MPA of skilled minorities from external source 1} = 90.$$

Similarly, the following equations are derived:

$$\begin{aligned} x_3 + x_8 &\leq \text{MPA of unskilled minorities from external source 1} = 200; \\ x_4 + x_9 &\leq \text{MPA of skilled Caucasians from external source 1} = 140; \\ x_5 + x_{10} &\leq \text{MPA of unskilled Caucasians from external source 1} = 30; \end{aligned}$$

Vol. 4
1976

TABLE 1. MANPOWER PLANNING MODEL VARIABLES*

Employment levels	Division A				Division B			
	External source[1] (male)	External source[2] (female)	Internal source[1] (male)	Internal source[2] (female)	External source[1] (male)	External source[2] (female)	Internal source[1] (male)	Internal source[2] (female)
Skilled minorities	x_2	x_{12}	5	0	x_7	x_{17}	2	0
Unskilled minorities	x_3	x_{13}	0	0	x_8	x_{18}	0	0
Skilled Caucasians	x_4	x_{14}	195	0	x_9	x_{19}	118	0
Unskilled Caucasians	x_5	x_{15}	0	0	x_{10}	x_{20}	0	0
Total	x_1	x_{11}	x_{21}	x_{23}	x_6	x_{16}	x_{22}	x_{24}
Total after transition			x_{25}				x_{26}	

* Where x_n = number of employees in each specified category.

$x_{12} + x_{17} \leq$ MPA of skilled minorities from external source 2 = 5;
$x_{13} + x_{18} \leq$ MPA of unskilled minorities from external source 2 = 300;
$x_{14} + x_{19} \leq$ MPA of skilled Caucasians from external source 2 = 5;
$x_{15} + x_{20} \leq$ MPA of unskilled Caucasians from external source 2 = 600.

The values of the total number of individuals presently employed by the company are incorporated into the constraint equations of the model. Since these are constants at the time of the analysis, no variable is assigned to these values, and they are included in the model as constant values for the internal sources 1 and 2 for divisions A and B:

$$\begin{aligned} x_{21} &= 200; \\ x_{22} &= 120; \\ x_{23} &= 0; \\ x_{24} &= 0. \end{aligned}$$

To examine movement between the divisions in the organization, management can develop a transition matrix based upon past experience. Although these values are constants, treating the figures as variables would provide management with the opportunity to allow for model modification based upon internal labor movements [1, p. 121].

Vol. 4
1976

In the developed matrix, 70 per cent of the people employed in division A will remain in their present positions and 20 per cent will move to division B during the next period. The remaining 10 per cent will be leaving the system for various reasons, such as being fired, retiring, illness or quitting for a position with another company. Similarly, of the total number of people presently employed in division B, 80 per cent will remain in division B during the period, 10 per cent will move to division A, and the remaining 10 per cent will leave the system. Since it has been indicated that the machinery is more complex in division B and the average salary is higher, it can be assumed that employment in division B may be preferable to employment in division A. The information given in the transition matrix supports this assumption. The following two equations summarize this data:

$$x_{25} - 0{\cdot}7x_{21} - 0{\cdot}7x_{23} - 0{\cdot}1x_{22} - 0{\cdot}1x_{24} = 0$$

and

$$x_{26} - 0{\cdot}8x_{22} - 0{\cdot}8x_{24} - 0{\cdot}2x_{21} - 0{\cdot}2x_{23} = 0.$$

The next two constraint equations are production constraints. The company has determined the optimal level of manpower in each division to be reached by the beginning of the next period in order to have the capability of producing the level of output required for the company to attain its projected profit position. For division A the desired manpower level is 300, and for division B it is 190. The production constraint equations are shown below:

$x_1 + x_{11} + x_{25} + d_1^- - d_1^+ =$ manpower level for division A at start of period 2 = 300

and

$x_6 + x_{16} + x_{26} + d_2^- - d_2^+ =$ manpower level for division B at start of period 2 = 190.

There are two budget constraints imposed on this basic manpower planning model. Each division must try to operate within a minimum salary budget, thereby minimizing the total organizational budget for employees' salaries in the model. The model attempts to minimize the sum of the salaries times the number of employees in each category. Using division A as an example, the average wage for external unskilled, external skilled, and internal employees is \$6000, \$8000 and \$9000, respectively. For division B, the average wage levels are similarly \$7000, \$9000 and \$10,000. The higher salaries for the internal sources are caused by the influence over the internal labor market by the employees [5, pp. 64–90]. Given these values and constraints, the following equations can be derived:

$$\$6000\,(x_3 + x_5 + x_{13} + x_{15}) + \$8000\,(x_2 + x_4 + x_{12} + x_{14}) + \$9000\,(x_{25}) - d_3^+ = 0$$

$$\$7000\,(x_8 + x_{10} + x_{18} + x_{20}) + \$9000\,(x_7 + x_9 + x_{17} + x_{19}) + \$10{,}000\,(x_{26}) - d_4^+ = 0.$$

The final constraint equation for the basic model is concerned with the number of unskilled people employed in division B. The equation simply indicates that the total number of unskilled workers employed from external sources for division B will be greater than or equal to zero. The rationale for this constraint deals with the higher skill requirement in division B. This is shown below:

$$x_8 + x_{10} + x_{18} + x_{20} - d_5^+ = 0.$$

It is seen that minimizing the positive deviation of this expression from zero will cause the number of unskilled workers hired for division B to be minimized. This will be shown in the objective function.

Additional constraints imposed by the environment. In recent years, government agencies have placed the responsibility on the firm of determining whether underutilization of minorities and women exists within their organization. If underutilization does exist, the company feels it must take positive steps to alleviate the situation by employing minorities and women in proportion to their availability in the recruiting area of the company. A possible problem, the overlapping between minorities and women, e.g. the black female, was not directly incorporated into the model since the company was concerned only with the total levels of minorities and females.

From the information given concerning the manpower availability in the recruiting area of the company, it is determined that approx. 43 per cent of the

total available work force is minorities. This means that for the company to satisfy the government, the number of minorities that they employ from external sources must be greater than or equal to 43 per cent of the total number employed from external sources. This is shown below:

$$x_2 + x_3 + x_7 + x_8 + x_{12} + x_{13} + x_{17} + x_{18} - 0{\cdot}43x_1 - 0{\cdot}43x_6 - 0{\cdot}43x_{11} - 0{\cdot}43x_{16} - d_6^+ = 0.$$

In a similar fashion, approx. 67 per cent of the available manpower in the area is women. Because of a company desire to equate the ratio of women employed in the organization and the ratio of women in the organization's labor market area, the number of women employed must be greater than or equal to 67 per cent of the total number employed from external sources.

$$0{\cdot}33x_{11} + 0{\cdot}33x_{16} - 0{\cdot}67x_1 - 0{\cdot}67x_6 - d_7^+ = 0.$$

In addition to the above constraints, an equation is needed to ensure that the percentage of minorities hired would be distributed between the skilled and unskilled categories according to their proportion in the general population

$$0{\cdot}57x_3 + 0{\cdot}57x_8 + 0{\cdot}57x_{13} + 0{\cdot}57x_{18} - 0{\cdot}43x_5 - 0{\cdot}43x_{10} - 0{\cdot}43x_{15} - 0{\cdot}43x_{20} - d_8^+ = 0.$$

Vol. 4
1976

In meeting the environmental constraints imposed in the manpower model, it is likely that the firm will be forced to employ unskilled workers. In order for the company to utilize these unskilled workers, it will be necessary to use a training program where it is assumed that the costs associated with the operation of the training program must come from the salary budgets of each division. Based on previous company experience, every person who is successfully trained for division A will cost division B \$111.50. Similarly, every person who is successfully trained for division B will cost division B \$124.22.

It was stated that funds for the training program must be generated from the divisional salary budgets. Therefore, the minimization of each divisional salary budget must be modified for the second stage by changing the average salary constant for unskilled workers in each division. Division A would be \$6111.50 (\$6000 + \$111.50), and division B would be \$8124.22 (\$8000 + \$124.22).

$$\$6111.50(x_3 + x_5 + x_{13} + x_{15}) + \$8000(x_2 + x_4 + x_{12} + x_{14}) + \$9000(x_{25}) - d_3^+ = 0,$$

$$\$8124.22(x_8 + x_{10} + x_{18} + x_{20}) + \$9000(x_7 + x_9 + x_{17} + x_{19}) + \$10{,}000(x_{26}) - d_4^+ = 0.$$

Objective function

The objective function of the company's manpower decision-making model is comprised of five goals or priorities which the model attempts to satisfy in descending order after the initial definitional constraints have been met. For

each of these priorities, P_i equals one when the deviations on the i^{th} priority are being minimized, otherwise the value is zero. The first priority of the company is concerned with the production of a desired level of output. It requires that the company maintain a certain level of manpower which is capable of producing the desired output. If this manpower level is not maintained, the employees will be forced to work overtime, a situation the company wishes to avoid. The company, therefore, wants to employ at least the number of people deemed necessary to produce the desired level of output. This objective may be achieved if the negative deviation from the manpower level in each division is minimized. The term expressing this in the objective function is shown below:

$$P_1(d_1^- + d_2^-).$$

The company's second objective is to avoid overemployment in the divisions. It wants to employ the fewest number of workers possible while still producing its desired level of output. The company is trying to maximize its production efficiency by effectively utilizing its manpower resources. The term shown below expresses the company's objective to avoid overemployment:

$$P_2(d_1^+ + d_2^+).$$

It was mentioned earlier that when an employee is hired, he may be placed in either division. It should be noted, however, that a new employee would usually start work in division A because the machinery in this division is more easily operated and the average salary is lower. The company's third priority is aimed at avoiding overemployment of unskilled employees if they are initially placed in division B. The term below minimizes the number of unskilled employees starting in division B:

$$P_3(d_5^+).$$

The fourth objective is to employ the desired number of minority members and women in the organization. Obviously, this attempted level would be based upon the distribution in the firm's recruiting area.

$$P_4(d_6^+ + d_7^+ + d_8^+).$$

The fifth priority is concerned with overruns in the total salary of each division. This objective will be attained by minimizing the positive deviation in each divisional salary expenditure. This is expressed by the term shown below:

$$P_5(d_3^+ + d_4^+).$$

When all of these company objectives are taken together, the result is the objective function of the manpower decision-making model, which is:

$$Z = P_1(d_1^- + d_2^-) + P_2(d_1^+ + d_2^+) + P_3(d_5^+) + P_4(d_6^+ + d_7^+ + d_8^+) + P_5(d_3^+ + d_4^+).$$

RESULTS

At this point, two successive computer runs of the manpower decision-making model were instituted. These two runs are based on the two stages in the development of the model. The results are presented in terms of the values that the variables attained in each run, as well as in terms of the degree of achievement of each priority.

In the first run, the firm desired to hire a certain percentage of minority group members and women in an attempt to utilize these people in proportion to their availability in the environment. This was a constraint in the manpower model brought about by environmental forces.

With these constraints and after rounding the results, the manpower decision-making model indicated that the firm should employ 13 unskilled, male minority group members (x_3); 51 unskilled, female minority group members (x_{13}); and 84 unskilled, female Caucasians (x_{15}) in division A at the start of period 2. These additional employees, plus the 152 employees remaining after transition (x_{25}), provide the necessary manpower level of 300 employees.

Vol. 4
1976

Similarly, the model forecasted that 31 skilled, male Caucasians (x_9) and 23 skilled, male minority members (x_7) should be employed in division B. This figure, plus the 136 employees remaining after all transitions (x_{26}), provides the necessary manpower level of 190 employees in division B at the start of period 2.

In run two, one additional factor was added to the model as the result of environmental forces requiring the firm to hire unskilled employees in run one. This additional factor was a training program.

With the salary constraints of each division, the model indicates that the firm should employ 46 unskilled, female minority group members (x_{13}); 18 unskilled, male minority group members (x_3); and 84 unskilled, female Caucasians (x_{15}) in division A. These employees, plus the 152 remaining after all transitions (x_{25}), provide the necessary manpower level of 300 employees in division A at the start of period 2.

Similarly, the model suggests that 23 skilled, minority group members (x_7); 26 skilled, male Caucasians (x_9); and 5 skilled, female Caucasians (x_{19}) be employed in division B. These employees, plus the 136 remaining after all transitions (x_{26}), provide the necessary manpower level of 190 employees in division B at the start of period 2.

From the priority analysis in the results, the model indicates that for both runs the needed manpower level for production was met, overemployment was avoided, no unskilled employees began work in division B, and sufficient levels of minority members and women were employed. Given the achievement of the above objectives, the total salary cost was minimized to \$4,101,995 without the training program and \$4,118,492 with the training program.

SUMMARY AND IMPLICATIONS

Hopefully, this work has demonstrated the need for firms to have definitive manpower information systems and methods and models of manpower planning. The point is that manpower planning must become more sophisticated if the firm hopes to be successful in acquiring necessary manpower given various constraints.

A generalized model was developed for use in the decision-making stage of the manpower planning process. The model can be utilized with any data that the firm can generate to fit the equations, and new equations can be developed for various firms. Obviously, many factors may occur which would require modification in the model. These problems could range from required logistic change to automation which require changes in the skill mix of manpower. Additionally, the environmental change process may be so uncertain because of existing turbulence that the required model has to be continually modified. This would occur because of the lack of a perfected method of deterministic goal programming.

As previously indicated, some positive characteristics of goal programming deal with its capability to construct models with a vast array of factors. These concepts could consist of modification of the transition matrix representing shifts in promotion or retirement policies for internal shifts between divisions. Furthermore, hiring and severance costs can be explicitly modeled, such as the incorporation of a training program.

Some negative factors about the model also exist. Because the priority rankings are subjectively determined, they may not be optimal for all executives in the organization. This could be alleviated by examining a number of various priority structures established by some form of team ranking which would hopefully approach the optimal structure for all managers in the firm. Additionally, to make the model realistic by incorporating all constraints, a great deal of time must be expended. Fortunately, this work would be extremely beneficial when tied with other organizational variables to provide executives with necessary decision-making information.

A final point to be made is that the result is only as good as the data; and even if the data are perfect, the result must be used in conjunction with the entire manpower planning process. Only in this way can the firm practice effective manpower planning.

REFERENCES

1. Bartholomew DJ & Smith AR (Eds.) (1970) *Manpower and Management Science.* The English Universities Press, London.
2. Bennis W (1966) *Beyond Bureaucracy.* McGraw-Hill, New York.
3. Burack EH (1972) *Strategies for Manpower Planning and Programming.* General Learning Corporation, Morristown, NJ.

4. BURACK EH, STASZAK FJ & PATI GC (1972) An organizational analysis of manpower issues in employing the disadvantaged. *Acad. Mgmt Jl* **15,** 255–271.
5. DOERINGER PB & PIORE MJ (1971) *Internal Labor Markets and Manpower Analysis.* DC Heath, Lexington, MA.
6. EMERY FE & TRIST EL (1965) The causal texture of organizational environments. *Hum. Relat.* **18,** 21–31.
7. GUETZKOV H, KOTLER P & SCHULTZ RL (1972) *Simulation in Social and Administrative Science.* Prentice-Hall, Englewood Cliffs, NJ.
8. LEE SM (1972) *Goal Programming for Decision Analysis.* Auerbach Publishers, Philadelphia, PA.
9. MCWHINNEY WH (1968) Organizational form, decision modalities, and the environment. *Hum. Relat.* **21,** 269–281.
10. SMITH AR (Ed.) (1961) *Models of Manpower Systems.* Elsevier, New York.
11. STAINER G (1971) *Manpower Planning: The Management of Human Resources.* Heinemann, London.
12. TERREBERRY S (1968) The evolution of organizational environments. *Admve Sci. Q.* **12,** 590–613.
13. The Conference Board (1971) *Manpower Planning; Evolving Systems.* The Conference Board, New York.
14. TOFFLER A (1970) *Future Shock.* Bantam, New York.
15. VETTER EW (1967) *Manpower Planning for High Talent Personnel.* The University of Michigan, Ann Arbor, MI.

ADDRESS FOR CORRESPONDENCE: *Professor H Kahalas, Department of Business Administration, Virginia Polytechnic Institute and State University College of Business, Blacksburg, VA 24061, USA.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Problems in the Application of Portfolio Selection Models

STEWART D HODGES

(*Received April 1975; in revised form April 1976*)

This paper discusses a number of problems which arise in the implementation of portfolio selection models. It is suggested that the effect of errors and biases in expected return forecasts can be reduced by using these forecasts to modify a prior distribution which leads to minimal trading activity. Efficient diversification across industry groups is hampered by the difficulties of predicting covariances. The use of a selection model through time raises the issues of revising forecasts and of the relationships upon which the appropriate investment horizon and portfolio turnover depend. Last, consideration is given to the conflicts which may exist between management objectives and the mean-variance criterion used by most models.

Vol. 4
1976

INTRODUCTION

A NUMBER of experiments have recently been made in the use of computerised models for managing portfolios of assets in a systematic way (see, for example, [1]). The interest in these methods stems from two main considerations. Firstly, it is hoped that they may lead to a reduction in the manpower and administrative costs required to manage funds, particularly in trust departments of banks which have responsibility for a large number of relatively small accounts. Secondly, they may provide a means of using the information available to a portfolio manager in the most effective possible way in terms of generating superior fund performance. Despite these potential advantages the use of such models is not currently widespread.

Portfolio selection models rely on a variety of data inputs including the subjective assessments of security analysts as to the likely returns from investments in individual common stocks. The implementation of such models needs to be based on knowledge of the statistical properties of these data, and on a clear definition of management objectives. The purpose of this paper is to discuss some conceptual and empirical problems which arise and must be dealt with in any implementation of portfolio selection models. Three problems areas are identified:

1. obtaining satisfactory inputs for a mean-variance selection model;

2. using a mean-variance model through time;
3. ensuring that the model is consistent with management objectives.

1. MEASUREMENT OF INPUTS FOR MEAN-VARIANCE MODELS

When the idea of mean-variance portfolio selection models was first introduced by Markowitz [13] computational difficulties were sufficient to make his approach impracticable for most portfolio managers. Since then however with successive simplifications by Sharpe [17, 19, 20] and dramatic progress in computing technology, the computational problems are now the very least of the difficulties facing the would-be user of these models.

The quality of forecasts of security returns

In an earlier simulation study [10], we indicated how the diagonal portfolio selection model (Sharpe [17]) might be used to manage a portfolio through time, taking into account the costs of making transactions and the limitations of ability to forecast common stock returns. Our approach resembled in some ways that of Treynor and Black [21] who dealt analytically with the concept of the investor's forecasting ability but who did so in a world with no transactions' costs. With these studies in mind it is apparent that knowledge of the extent of one's ability to forecast individual stock returns is crucial to the portfolio selection process. This implies that considerable effort should be spent on making analysts quantify their forecasts, and on performing analyses to obtain as much information as possible on the statistical properties of such forecasts.

Relatively few studies have been published which set out to measure the ability of analysts to forecast the returns on individual common stocks. In view of the tremendous amount of activity by academics in testing various other forms of the efficient markets hypothesis (see [9] for an excellent survey) this is perhaps surprising, but the information required is typically very confidential, and often it is not collected in a suitably quantified form. The studies by Ambachtsheer [1], and Diefenbach [7] are fairly typical in indicating low but apparently significant degrees of forecasting ability. Many similar studies done within other financial institutions have never been published for reasons of confidentiality.

In an efficient market, analysts would have no ability to forecast the price changes of stocks so as to earn abnormally high returns, since all available information would be incorporated in the stock prices at all times and only a fair return could be expected. In such a market, portfolio selection models would have very limited usefulness since trading should only be undertaken to maintain a diversified position and possibly to maintain some target level of systematic risk. The emphasis by academics on the efficient markets hypothesis helps to explain why comparatively little work has been done on the development of

portfolio selection models. On the other hand if analysts possess a significant degree of forecasting ability, or more to the point, if they are expected to possess such forecasting ability in the future, then portfolio turnover, composition and performance should depend directly on its extent. If analysts could forecast the returns on individual stocks very accurately then portfolio selection models would again be redundant, as it would then only be necessary to pick individual stocks instead of diversified portfolios. It is in the in-between world of a market which behaves fairly efficiently, but in which a professional analyst can forecast stock returns with some degree of success that portfolio selection models may have some usefulness in translating his forecasting ability into superior portfolio performance. The relatively poor performance of simple mechanical strategies and of people operating in competition with a diagonal model in simulated conditions lends support to this view (see [10] and [5] for studies of this kind).

The degree of forecasting ability can be measured as the correlation coefficient between the forecasts of common stock returns and their realisations, where both are adjusted for movement of the market. The difficulties of obtaining adequate information of this kind on the statistical quality of forecasts cannot be over-emphasized. The Hodges and Brealey study indicated the dangers of assuming different categories of forecasts (as say from different analysts or for different industry groups) were of the same quality when they were in fact different. The need to disaggregate to attempt to estimate systematic differences of this kind, and the limitations of time series data, mean that it will never be possible to estimate the statistical properties of forecasts with much precision. However, even if we are unable to estimate the appropriate correlation coefficient with much confidence, in order to measure the degree of forecasting ability of an analyst, it may be possible to develop selection models which can take its uncertainty into account.

Vol. 4
1976

Adjusting input data for selection models

The aim in estimating inputs must be to refine them until they have as nearly as possible the statistical properties required by the selection model. It should not be possible to devise a stratification which will reveal any systematic bias. For example, if the cross-sectional dispersion of forecast returns is estimated wrongly, the high and low forecasts will be biased in opposite directions resulting in either too much or too little turnover. Hodges and Brealey suggested the use of expected returns consistent with the capital asset pricing model [18] as Bayesian priors which are then modified by the analyst's forecasts according to his estimated forecasting ability. This procedure ensures that the model behaves correctly with low turnover for low forecasting ability. The concept can also be extended to more complex representations of security risk and its relationship to required rates of return (see [11] for a survey of pricing models). The use of such a method also minimizes the effect of estimation errors in the parameters

(described below) which are employed to characterize security and portfolio risk.

The diagonal model quantifies risk as the sum of two components: systematic risk due to market movements which cannot be reduced by diversification and unsystematic risk which can be diversified away. The systematic risk of a stock (the extent to which it has a tendency to move up and down with the market as a whole) is represented by its beta value. The beta of a stock is usually measured as the slope of a regression line between returns on the stock (as the dependent variable) and returns on a market index (as the independent variable). The betas of individual stocks are used to calculate the betas of portfolios. Blume [3] has shown how a cross-sectional regression between sets of betas estimated in different time periods can be used to adjust them for bias and empirically finds a need to adjust about 40% towards the mean. Vasicek [22] gives a model for the adjustment to each stock, based on using the cross-sectional distribution of betas as a Bayesian prior. The adjustments are of the same form as Blume's but empirically they appear on average to be slightly too small, presumably because of a combination of specification problems regarding the shape and stationarity of the probability distributions, and the stationarity of the betas (see also [4]).

The unsystematic risk of a stock is quantified by the estimated residual variance around the regression line employed to measure beta. These variances may be adjusted in a similar way, though in this context it may be less important; errors here principally affect how much of a stock to buy, rather than which stock, and should have a smaller impact on total portfolio risk. The proportion by which an estimated variance is adjusted by the Bayesian method towards the average variance depends on the size of the estimated variance itself. If the estimate is large the proportion can be very significant.

The variance–covariance structure

The structure of the variances and covariances between stock returns deserves some attention. In a world where many managers believe that they are better at picking industries than at picking individual stocks within industries, it might be tempting to concentrate in a few particular industry sectors.

Since the yardsticks for comparison across industries are much less sure than within any particular industry, a manager may plausibly expect to earn above average returns from concentrating in particular sectors on the basis of superior economic analysis, but these opportunities are limited by the extent to which industrywide movements exist and can be forecast. Multi-index or full Markowitz models are necessary for a satisfactory normative solution to the problem of diversification across industry sectors, and the diagonal model may be a less good approximation than had previously been supposed.

It is worth pointing out at this point a common fallacy involving portfolio selection models which bears on this. In comparing his diagonal model with the

Markowitz one, Sharpe [17] used an average of ex post realised returns as a proxy for expectations. The variance structure was also estimated directly from historical time series. Many others have also computed efficient portfolios in this way including Cohen and Pogue [6] in their assessment of their multi-index models and a number of authors in studies of international diversification. In the absence of suitable subjective data the approach seems a plausible one, but averages of the past generally have too much dispersion to be plausible as naive expectations of the future and they may imply a degree of forecasting ability significantly greater than that intended by the authors. The spread of expected returns should always be less than that of the ex post cross-sectional distribution corresponding to the same time horizon. The ratio of these spreads should approximate to the correlation coefficient between forecasts and outcomes.

As this correlation increases, the cross-sectional distribution of expectations spreads outwards towards the values which will be realised, and at the same time the variances about those expectations decrease. The mixing of ex-post returns as expectations (giving some degree of spread and implying significant forecasting ability) with historically estimated variances is clearly incorrect. The usual effect is for it to imply unrealistically high forecasting ability, despite the variances now being exaggerated, and this renders comparisons rather meaningless.

Vol. 4
1976

The factor analysis approach of King [12] and more recently Meyers [14] is also of limited value for portfolio selection. While industry factors (and other common factors besides the market factor) are clearly significant, and should be taken into account in portfolio diversification, knowledge of the nature and stability of the correlations between these factors is missing. The recent empirical study by Elton and Gruber [8] on forecasting the correlation matrix of stocks is a welcome step forward, but this did not include any Bayesian methods of the kind described earlier and which in principle would lead to improved forecasts.

2. USING PORTFOLIO MODELS THROUGH TIME

A number of problems arise when using a portfolio selection model to manage funds through time. Ideally one would like to use some form of multi-period model rather than a single horizon one. This can give a proper basis for comparison of investments which are expected to take different lengths of time to recoup their transaction' costs and start earning excess returns, and also attacks the problem of the best timing of transactions. Models of this kind, Pogue [15] for example, do perform this type of analysis, but it is inevitably at the expense of demanding a much greater range of inputs on forecast returns, and the difficulties of obtaining inputs with known statistical properties may be too great to make them worthwhile.

Both single and multi-period models are likely to be inefficient if there is any

random noise changing the forecasts through time and causing additional portfolio changes before the final horizon period used by the model in its original selection is reached. The single period selection model can be viewed in terms of two cross-sectional distributions of expected excess returns. One distribution is for the set of stocks which are already held in the portfolio; the second distribution is for the (generally larger) population of stocks which are eligible to be bought for the portfolio. The selection model sells stocks in the left tail of the distribution of stocks held and buys ones in the right tail of the distribution of eligible stocks until (for the marginal transaction) the difference of expected return between the best eligible stock and the worst stock held equates to the level of transactions' costs. If there is random noise affecting these expectations, revisions before the horizon period is reached will partly reflect that noise and will be detrimental to portfolio performance. Again, more empirical evidence is required on the statistical properties of forecasts which are revised through time before models can be developed to exploit them efficiently and make portfolio decisions on a continuing basis.

Viewing the selection model in terms of these two distributions of expected returns yields some information on the levels of turnover and performance which can be expected as a function of forecasting ability, transactions' costs and the choice of horizon period. As the portfolio is held through time, its cross-sectional distribution of expected excess returns will spread out and shift to the left as these returns are realised until it becomes representative of the total universe of stocks. Just how fast this occurs is an empiricial question, but in some cases it may be reasonable to suppose that by the end of the horizon period the two distributions are already very similar, if the original forecasts have little or no predictive content beyond the horizon date. This assumption will be used to illustrate the approach.

Assume that the portfolio consists of large number of equal sized blocks of stock and that there is also a large population of stocks eligible to be bought. Figure 1 illustrates the distributions of expected excess returns for the two sets of stocks. The standard deviation (S) of both of these distributions is the product of three terms. These are the annual residual standard deviation (σ) of a typical stock (removing systematic variation with the market), the correlation coefficient (ρ) between forecast and actual excess returns, and the square root of the time to the horizon,

$$S = \sigma\rho\sqrt{t}. \tag{1}$$

All stocks held in the portfolio which have expected excess returns less than some value R_S will be sold and eligible stocks with expected excess returns greater than some value R_B will be bought. Since transactions will be made up to the marginal point at which they are expected to only just recoup the cost of making them, R_S and R_B are related to the level of transactions' costs C by the equation

$$R_B = R_S + C. \tag{2}$$

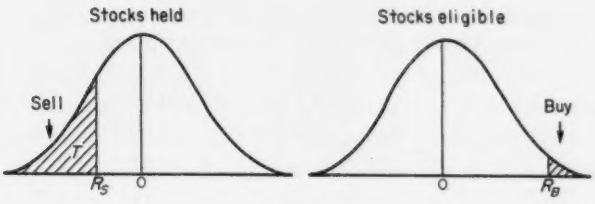


FIG. 1.

The shaded area on the left hand side of Fig. 1 represents the stocks sold and gives the turnover per horizon period (T). The right-hand shaded area represents the stocks bought and is smaller than the left-hand area in the ratio (n) of the number of stocks currently held to the number eligible.

Thus using the notation $Pr\{\}$ to denote probability, the values of R_S and R_B are uniquely determined by (2) and the relationship

$$Pr\{R < R_S\} = nPr\{R > R_S + C\}. \tag{3}$$

R_S and R_B can be computed numerically for any known probability distribution.

The expected excess return per horizon period, R_t, is then given by integrating across the two shaded areas as

$$R_t = \int_{-100}^{R_S} (R_S - x)f(x)\mathrm{d}x + n\int_{R_B}^{\infty} (x - R_B)f(x)\mathrm{d}x \tag{4}$$

where $f(x)$ denotes the density function of the distribution of percentage excess returns, and the integration runs from the lowest possible return of -100%. This approach can be used to explore the relationships between forecasting ability and portfolio turnover and performance. The expected annual excess return and probable portfolio turnover can be tabulated for different choices of horizon length as a function of the level of forecasting ability and the number of stocks analyzed, assuming estimated values for the level of transactions' costs and the annual residual variances of the stocks. This can assist in the choice of a suitable length for the horizon period.

3. CONSISTENCY WITH MANAGEMENT OBJECTIVES

If management is willing to accept the portfolios selected by a model then its objectives are probably consistent with those of management. The converse does not hold, for portfolios which management regard as inappropriate may be a result of an inadequate specification elsewhere, or due to measurement errors of the kinds discussed above. One way around this problem is to impose

Vol. 4
1976

additional constraints which will make the portfolios acceptable. Sharpe (20) describes these types of constraints and he points out the advantages of using a linear programming method for incorporating them. Constraints on current yield, the realization of capital gains, the amount of portfolio turnover, and the size of holding in particular industry groups may all be added in this way. The problem of budgeting for estimated outflows at different future dates can also be handled by adding linear constraints. Constraints of this kind represent at best an *ad hoc* compromise and at worst dangerous suboptimization, but do appear to be necessary given the current state of the art.

More serious difficulties arise in terms of the consistency of the conventional mean-variance approach with management objectives. With transactions' costs included, and a suitably conservative idea of the level of forecasting ability, the portfolios selected by the diagonal model are fairly well diversified. Nevertheless, they may still have a much greater degree of residual risk than many managements can tolerate. The issue is really whether the objectives of management are to select portfolios which are efficient in nominal terms, in real terms, or relative to some other portfolio such as the market index, or some combination of all of these.

Most work so far has been done in nominal terms, and considerable empirical and theoretical work remains to be done before it will be possible to proceed comfortably in real terms. Pye (16) has suggested that the objective of minimising the maximum regret has some applicability. While there are some problems with this particular criterion (which is generally inconsistent with the Von Neuman–Morgenstern axioms) it is certainly true that most managers are assessed not by their absolute level of performance but relative to what the market has done. The conventional mean-variance approach obtains the maximum expected return for what ever level of variance of return is accepted. If performance relative to the market index (or some other yardstick) is what matters, then the variance of the difference in return from this index may be controlled rather than the overall variance of return. Instead of calculating efficient portfolios in the mean-variance space of the return on the portfolio (R_p), the mean-variance space of the difference between the portfolio and market returns ($R_p - R_m$) can be used. In the diagonal model this is done by modifying the equation which gives the sensitivity of the portfolio return to a change in the market index (its beta), to give instead the sensitivity of the difference between the portfolio and market index returns (beta minus one). While not necessarily to be advocated this can be consistent with the axioms (as the maximisation of the expected value of a state dependent utility function) over some classes of probability distributions.

Figure 2 illustrates efficient portfolios under these alternative views of risk. They were calculated using simulated data for a population of 50 stocks, with transactions' costs at 4 per cent, a 1-year horizon period and a correlation coefficient of 0·10 between forecast and actual returns. Figure 2 gives the usual

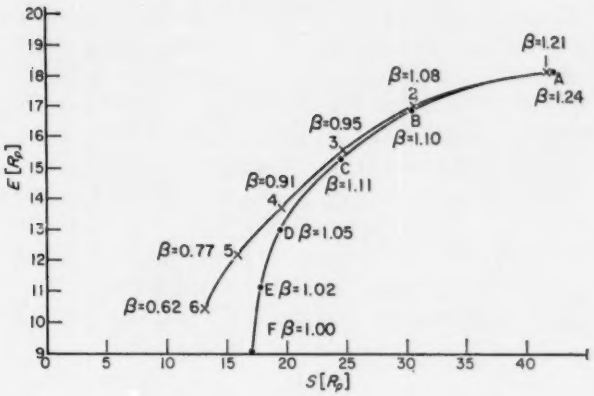


FIG. 2.

mean — standard deviation space ($E - S$) of portfolio returns, R_p; the portfolios numbered 1–6 are efficient in this space as they provide the maximum possible expected return for each level of standard deviation of return. Below this efficient frontier is a second curve of portfolios (including the portfolios denoted by the letters *A–F*) which give somewhat lower expected returns for a given standard deviation of return. However, these portfolios are efficient in terms of their differences from the market return: they give the highest expected returns for any given standard deviation of the difference between market and portfolio returns. Portfolios *A–F* are on the efficient frontier in this space and dominate the portfolios (1–6) which were efficient in the usual mean — standard deviation space of Fig. 2. The market sensitivities (β's) of these portfolios are also indicated in Fig. 2.

It is no surprise that each efficient frontier dominates the other in its own space, but the extent of the domination is quite striking. Note particularly that none of the usual efficient portfolios has the standard deviation of the difference between its return relative to the index of less than 10 per cent, so it always has a large chance of being beaten by the market. In contrast we can find a portfolio with a standard deviation of 5 per cent relative to the index with an expected return over 2 per cent above that of the index. The portfolios selected are also quite different along the two frontiers. The portfolios which are efficient relative to the index tend to have betas closer to one and are more diversified than the usual efficient portfolios. Not suprisingly they also involve less turnover through time.

Both the differences between the frontiers and between the compositions of the portfolios composing them depend on the chosen values of the inputs described above, but the results illustrate clearly the conflicts involved between different plausible objectives. These are compounded when consideration is given to real returns rather than just to nominal returns. The definition of an acceptable form of objective seems considerably more complicated than the

Vol. 4
1976

advocates of mean-variance analysis have led us to believe. Even supposing that viewing portfolio returns relative to the market has no relevance for the investor, it may still be important to the investment manager as it affects his compensation and job security in a competitive environment. This raises the issue of formulating appropriate incentive structures for management. Some particularly awkward problems are raised when consideration is given to setting objectives in terms of real purchasing power, since almost all existing statistical work is in nominal terms and the choice of a price index deflator will always be to some extent arbitrary. A recent paper by Biger (2) gives a comparison of real and nominal returns in portfolio selection but unfortunately falls prey to the fallacy described earlier of using averages of the past to proxy as expectations.

SUMMARY

The paper has discussed a number of problems which arise in the implementation of normative models for portfolio selection. Some methods of reducing biases in the inputs were outlined and the use of a neutral prior distribution, leading to minimal trading activity, modified by analysts forecasts was suggested. A further difficulty in diversifying efficiently across industry sectors lay in forecasting the variance–covariance matrix of returns. The use of a selection model through time raised the dual problems of securities with different expected holding periods and of the effect of revisions of forecasts through time. A simplified model of the selection process was developed to explore the relationships between forecasting ability, the number of stocks analysed, the length of the horizon period and the level of transactions' costs and portfolio performance and turnover. Finally the issue of conflicting management objectives was raised. It was shown how the diagonal model could easily be adapted to generate portfolios which are efficient relative to the market index. The manager's dilemma was posed as one of placing appropriate weights on obtaining efficient nominal returns, on obtaining efficient real returns and on reducing his risk of bad performance relative to the market.

REFERENCES

1. Ambachtsheer K (1974) Profit potential in an 'almost efficient' market. *J. Portfolio Mgmt* **1**, 84.
2. Biger N (1975) The assessment of inflation and portfolio selection. *J. Fin.* **30**, 451.
3. Blume ME (1971) On the assessment of risk. *J. Fin.* **26**, 1.
4. Blume ME (1975) Betas and their regression tendencies. *J. Fin.* **30**, 785.
5. Brealey RA & Hodges SD (1975) Playing with portfolios. *J. Fin.* **30**, 125.
6. Cohen KJ & Pogue GA (1967) An empirical evaluation of alternative portfolio selection models. *J. Bus.* **40**, 166.

7. DIEFENBACH RE (1972) How good is institutional brokerage research? *Fin. Analysts J.* **28**, 54.
8. ELTON EJ & GRUBER MJ (1973) Estimating the dependence structure of share prices. *J. Fin.* **28**, 1203.
9. FAMA EF (1970) Efficient capital markets: a review of theory and empirical work. *J. Fin.* **25**, 383.
10. HODGES SD & BREALEY RA (1973) Portfolio selection in a dynamic uncertain world. *Fin. Analysts J.* **29**, 50.
11. JENSEN MC (1972) Capital markets: theory and evidence. *Bell J.* **3**, 357.
12. KING BF (1966) Market and industry factors in stock price behavior. *J. Bus.* **39**, 139.
13. MARKOWITZ HM (1952) Portfolio selection. *J. Fin.* **7**, 77.
14. MEYERS SL (1973) A re-examination of market and industry factors in stock price behavior. *J. Fin.* **28**, 695.
15. POGUE GA (1970) An inter-temporal model for investment management. *J. Bank Res.* **1**, 17.
16. PYE G (1970) A note on diversification. *J. Fin. Quant. Anal.* **9**, 131.
17. SHARPE WF (1963) A simplified model for portfolio selection. *Mgmt. Sci.* **9**, 277.
18. SHARPE WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Fin.* **19**, 425.
19. SHARPE WF (1967) A linear programming algorithm for mutual fund portfolio selection. *Mgmt Sci.* **13**, 499.
20. SHARPE WF (1971) A linear programming approximation for the general portfolio analysis problem. *J. Fin. Quant. Anal.* **6**, 1263.
21. TREYNOR JL & BLACK F (1973) Using security analysis to improve portfolio selection. *J. Bus.* **46**, 66.
22. VASICEK OA (1973) A note on using cross-sectional information in Bayesian estimates of security betas. *J. Fin.* **28**, 1233.

ADDRESS FOR CORRESPONDENCE: *Dr. SD Hodges, London Graduate School of Business Studies, Sussex Place, Regent's Park, London NW1 4SA, UK.*

Vol. 4
1976

BLANK
PAGE

Vol.
197

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Multi-Terminal Vehicle-Dispatch Algorithm

BILLY E GILLETT

JERRY G JOHNSON

(*Received October 1975; in revised form June 1976*)

This paper introduces the Multi-Terminal Sweep Algorithm, a heuristic algorithm for obtaining an approximate solution to the multiple terminal vehicle-dispatch problem. The procedure determines a set of routes by which vehicles from two or more terminals can service a collection of demand points so that the total distance traveled is kept near to the minimum. This solution also satisfies constraints on the vehicle load and on the length of each route. Application of the algorithm to eleven multiple terminal vehicle-dispatch problems shows that near-optimal solutions to large-scale problems can be found in a reasonable amount of computer time.

Vol. 4
1976

INTRODUCTION

THIS PAPER considers the following type of problem. A fleet of vehicles stationed at M terminals is to deliver specified quantities of a single type of product to N locations in such a manner that the total distance traveled by the vehicles is minimized. Each vehicle has the same load capacity C and the same maximum range D. Each demand point occupies a unique location and has a positive demand for goods that can be satisfied by one visit of a single vehicle. There is no limit on the number of vehicles used or the quantity of goods dispatched from a terminal.

A variety of heuristic approaches have been introduced for the solution of single-terminal vehicle-dispatch problems, but only those by Tillman and Cain [6], Wren and Holliday [7], and Golden, Magnanti and Nguyen [5] have been capable of solving a problem with more than one terminal.

THE MULTI-TERMINAL SWEEP ALGORITHM

The Multi-Terminal Sweep Algorithm employs the strategy of partitioning the multiple-terminal vehicle-dispatch problem into a collection of subproblems

and solving each subproblem as a single-terminal problem. This approach reduces the amount of computation time required for obtaining a near-optimal solution by reducing the degree of complexity involved in solving each subproblem. However, an acceptable near-optimal solution usually requires that the procedure investigate a sequence of refinements or adjustments to the initial solution, because the partitioning process does not guarantee an optimal partition of the multi-terminal problem.

The requirement for the solution of many subproblems necessitates the use of a highly efficient single-terminal vehicle-dispatch algorithm. A modified version of the Sweep Algorithm, a vehicle-dispatch algorithm developed by Gillett and Miller [3], was chosen for this task. The Modified Sweep Algorithm, developed by Gillett and Johnson [4], exhibits computation times that are a linear function of the number of demand points when one considers problems having the same average number of locations per route. This property enables the Modified Sweep Algorithm to solve many types of large-scale vehicle-dispatch problems in reasonable amounts of time.

Since the Modified Sweep Algorithm is capable of producing very acceptable near-optimal solutions to single-terminal problems, the proper partitioning of demand points into clusters about individual terminals is the key to the solving of multiple-terminal problems. The optimal clustering pattern is a function of the spatial distribution of the demand points and of the load and distance constraints on the routes. The Multi-Terminal Sweep Algorithm places its principal emphasis on the spatial distribution criterion and attempts to assign demand points to terminals so that compact, widely separated clusters are formed.

Each cluster is composed of a single chain of demand points which encircles a single terminal. It is not intended that these chains of demand points represent feasible routes, but rather that they represent the arrangement of points along the periphery of the clusters. The clusters are built up by inserting unassigned demand points, one at a time, into the partially formed clusters. The Multi-Terminal Sweep Algorithm attempts to take into consideration the distance separating the clusters by minimizing, for each demand point, the extra distance that a vehicle would have to travel to include that point in the periphery about the selected terminal. This insertion cost, or the extra distance needed to include point j in the chain between points i and k, is calculated from the expression $d(i,j) + d(j,k) - d(i,k)$, where d represents the symmetric distance matrix for the problem.

The demand points are considered for assignment to terminals, so that the clusters are initially formed very close to the terminals and gradually expand outward. The minimum insertion cost is the criterion for determining the terminal to service each demand point. However, in order to resolve the assignment question in a region which is equidistant from two clusters, a second criterion is employed. If j is the demand point to be considered for assignment,

the algorithm determines the unassigned demand point k which is nearest to it. An estimate of the cost of assigning both j and k to each of the two nearest clusters is then calculated. If the terminal with the minimum insertion cost for j alone also has a smaller estimated insertion cost for the two points j and k, then j is assigned to that terminal. Otherwise, the assignment of location j is deferred until after point k has again been considered for assignment. During this cluster forming process the distance constraint must be checked so that no demand point will be assigned to a terminal too distant to service it.

The initial solution, obtained by solving the single-terminal vehicle-dispatch problems defined by the initial assignment procedure, may be capable of further improvement. It is possible that the reassignment of several demand points from one terminal to another might permit a significant reduction in travel distance at one terminal, so that an overall improvement results. In order to maintain the principle of compact, widely separated clusters, a modification of the initial assignment procedure is used to perform the reassignment of demand points.

The procedure selects for possible reassignment a set of demand points that are situated in the regions between the terminals. Each terminal takes a turn as an 'attracting center' and attempts to draw these demand points away from adjacent terminals. The demand point which is a candidate for reassignment, as well as some of its neighbors in the same cluster, is declared to be unassigned and the cluster forming process begins again. If the candidate location is reassigned to the fully formed cluster about the attracting terminal, the assignment procedure continues to completion and two single-terminal vehicle-dispatch problems are solved again. If an improved solution results, the changes become part of the current best solution, and this becomes the basis for further refinement. Demand points whose reassignment results in an improved solution are removed from the list of candidates for further reassignment. Some candidate points whose reassignments do not show much promise of yielding any improvements are omitted from consideration. Candidate points are considered in an order that is determined by their bearing angles with respect to the attracting terminal. The solution which results after all terminals have served as attracting centers is a near-optimal solution to the multiple-terminal vehicle-dispatch problem.

COMPUTATIONAL RESULTS

The Multi-Terminal Sweep Algorithm was used to solve eleven multiple terminal vehicle-dispatch problems. These problems were derived from single-terminal problems found in the literature [1, 3].

Since none of the published papers that deal with multiple terminal procedures provided details of the specific problems that were solved, no exact comparison

Vol. 4
1976

is possible. Tillman and Cain [6] provide some results for two problems with five terminals and fifty demand points. Their problems involved an average of 4·3 locations per route and an average execution time of 6·51 min on an IBM 360/50. This execution time can be compared with that of problem one, a four-terminal fifty-location problem with an average of 5·17 locations per route. Since the IBM 370/168 is roughly ten to fifteen times faster than the IBM 360/50, the 9·34 sec required to solve problem one is equivalent to about 1·5–2·4 min of IBM 360/50 time. Consequently, the multi-terminal procedure described in this paper is definitely competitive with Tillman and Cain's approach.

One important fact that emerges from the computational results is the relative insensitivity of computation time to the number of terminals in a multiple terminal problem. While a two-terminal problem requires considerably more time than a similar single-terminal problem, the computation time usually decreases as the number of terminals increase from two. Thus, all other parameters remaining the same, there appears to be no serious restriction associated with solving problems with a large number of terminals.

Problems with two terminals require a great deal of computation time because each refinement involves applying the Modified Sweep Algorithm to all N demand points. As more terminals are included in the problem, only those locations assigned to the two terminals associated with a given refinement are involved in the calculations. Thus, the more terminals in the problem the larger is the portion of locations that are not involved in the calculations. On the other hand, an increase in the number of terminals may result in an increase in the number of candidates for reassignment, and hence in an increase in execution time. The net effect observed in the eleven problems solved is that an increase in the number of terminals usually decreases the computation time.

Vol. 197

CONCLUSIONS

The Multi-Terminal Sweep Algorithm has the following attributes:

(a) It is a new approach to the solving of the multi-terminal vehicle-dispatch problem, and can handle problems with many terminals and hundreds of demand points in a reasonable amount of time.

(b) When the average number of locations per route and the total number of locations remain constant, an increase in the number of terminals above two usually decreases the computation time.

(c) As with the single-terminal Sweep Algorithm, the computation time increases quadratically with an increase in the average number of locations per route as long as the total number of locations remains constant.

(d) If the number of terminals and the average number of locations per route remain constant, the computation time increases in proportion to the square of the total number of locations.

(e) The computation time required by the Multi-Terminal Sweep Algorithm compares favorably with that published for other multi-terminal procedures.

TABLE 1. LIST OF PROBLEMS

Problem No.	Source of parent problem	Number of terminals	Number of demand pt.	Maximum load	Maximum distance
1	Christofides and Eilon[1]	4	50	80	Unlimited
2	Christofides and Eilon[1]	4	50	160	Unlimited
3	Christofides and Eilon[1]	5	75	140	Unlimited
4	Christofides and Eilon[1]	2	100	100	Unlimited
5	Christofides and Eilon[1]	2	100	200	Unlimited
6	Christofides and Eilon[1]	3	100	100	Unlimited
7	Christofides and Eilon[1]	4	100	100	Unlimited
8	Gillett and Miller[3]	2	249	500	310
9	Gillett and Miller[3]	3	249	500	310
10	Gillett and Miller[3]	4	249	500	310
11	Gillett and Miller[3]	5	249	500	310

TABLE 2. SOLUTIONS OBTAINED BY THE MULTI-TERMINAL SWEEP ALGORITHM

Problem No.	Average No. of locations/route[1]	Number of iterations[2]	Solution	CPU sec IBM 370/168
1	5·17	1	593·1606	9·34
2	9·33	12	486·1902	13·23
3	7·82	24	652·3828	20·43
4	7·19	8	1066·651	51·24
5	13·50	10	778·8760	227·67
6	7·25	21	912·2253	53·05
7	7·25	25	939·4587	52·48
8	10·58	17	4832·039	457·61
9	10·58	26	4219·691	444·33
10	10·58	41	3821·981	344·04
11	10·22	48	3754·147	315·99

[1] Includes terminal.
[2] The initial solution and any refinement which utilized the Sweep Algorithm count as an iteration.

REFERENCES

1. CHRISTOFIDES N and EILON S (1969) An algorithm for the vehicle-dispatching problem. *Opl Res. Q.* **20,** 309–318.
2. DANTZIG GB and RAMSER JH (1959) The truck dispatching problem. *Ops. Res.* **12,** 80–91.
3. GILLETT BE and MILLER LE (1974) A heuristic algorithm for the vehicle-dispatch problem. *Ops. Res.* **22,** 340–349.
4. GILLETT BE and JOHNSON JG (1974) Sweep algorithm for the multiple depot vehicle-dispatch problem. 46th National Meeting of ORSA, San Juan, Puerto Rico.

Vol. 4
1976

5. GOLDEN BL, MAGNANTI TL and NGUYEN HQ (1975) Implementing vehicle routing algorithms. Tech. Report 115, O.R. Center, MIT.
6. TILLMAN FA and CAIN TM (1972) An upperbound algorithm for the single and multiple terminal delivery problem. *Mgmt Sci.* **18,** 664–682.
7. WREN A and HOLLIDAY A (1972) Computer scheduling of vehicles from one or more depots to a number of delivery points. *Opl. Res. Q.* **23,** 333–344.

ADDRESS FOR CORRESPONDENCE: *Dr. Billy E Gillett, Department of Computer Science, University of Missouri-Rolla, Rolla, MI 65401, USA.*

APPENDIX A

The data for problems 8–11 in Table 1 and the corresponding route for problem 11 using the Multi-Terminal Sweep Algorithm are given in Tables 3 and 4.

TABLE 3. PROBLEMS 8–11 OF TABLE 1

Problem	Depots	Depot coordinates
8	2	(−33, 33), (33, −33)
9	3	(70, 0), (−50, 60), (−50, −60)
10	4	(75, 0), (0, 75), (−75, 0), (0, −75)
11	5	(70, 0), (40, 80), (40, −80), (−60, 20), (−60, −20)

Loc. No.	*X*	*Y*	Demand	Loc. No.	*X*	*Y*	Demand	Loc. No.	*X*	*Y*	Demand
1	−99	−97	6	29	79	38	8	57	−28	73	100
2	−59	50	72	30	−62	− 2	24	58	−76	55	62
3	0	14	93	31	−90	−68	53	59	41	42	90
4	−17	−66	28	32	52	66	13	60	92	40	57
5	−69	−19	5	33	−54	−50	47	61	−84	−29	44
6	31	12	43	34	8	−84	57	62	−12	42	37
7	5	−41	1	35	37	−90	9	63	51	−45	80
8	−12	10	36	36	−83	49	74	64	−37	46	60
9	−64	70	53	37	35	− 1	83	65	−97	35	95
10	−12	85	63	38	7	59	96	66	14	89	56
11	−18	64	25	39	12	48	42	67	60	58	56
12	−77	−16	50	40	57	95	80	68	−63	−75	9
13	−53	88	57	41	92	28	22	69	−18	34	39
14	83	−24	1	42	− 3	97	56	70	−46	−82	15
15	24	41	66	43	− 7	52	43	71	−86	−79	4
16	17	21	37	44	42	−15	12	72	−43	−30	58
17	42	96	51	45	77	−43	73	73	−44	7	73
18	−65	0	47	46	59	−49	32	74	− 3	−20	5
19	−47	−26	88	47	25	91	8	75	36	41	12
20	85	36	75	48	69	−19	79	76	−30	−94	3
21	−35	−54	48	49	−82	−14	79	77	79	−62	8
22	54	−21	40	50	74	−70	4	78	51	70	31
23	64	−17	8	51	69	59	14	79	−61	−26	48
24	55	89	69	52	29	33	17	80	6	94	3
25	17	−25	93	53	−97	9	19	81	−19	−62	52
26	−61	66	29	54	−58	9	44	82	−20	51	99
27	−61	26	5	55	28	93	5	83	−81	37	29
28	17	−72	53	56	7	73	37	84	7	31	12

TABLE 3—*contd*

Loc. No.	X	Y	Demand	Loc. No.	X	Y	Demand	Loc. No.	X	Y	Demand
85	52	12	50	140	− 5	−39	55	195	−55	39	19
86	83	−91	98	141	8	−22	45	196	70	−14	90
87	− 7	−92	4	142	−61	−76	100	197	0	95	35
88	82	−74	56	143	76	−22	38	198	−45	7	76
89	−70	85	24	144	49	−71	11	199	38	−24	3
90	−83	−30	33	145	−30	−68	82	200	50	−37	11
91	71	−61	45	146	1	34	50	201	59	71	98
92	85	11	98	147	77	79	39	202	−73	−96	92
93	66	−48	4	148	−58	64	6	203	−29	72	1
94	78	−87	36	149	82	−97	87	204	−47	12	2
95	9	−79	72	150	−80	55	83	205	−88	−61	63
96	−36	4	26	151	81	−86	22	206	−88	36	57
97	66	39	71	152	39	−49	24	207	−46	− 3	50
98	92	−17	84	153	−67	72	69	208	26	−37	19
99	−46	−79	21	154	−25	−89	97	209	−39	−67	24
100	−30	−63	99	155	−44	−95	65	210	92	27	14
101	−42	63	33	156	32	−68	97	211	−80	−31	18
102	20	42	84	157	−17	49	79	212	93	−50	77
103	15	98	74	158	93	49	79	213	−20	− 5	28
104	1	−17	93	159	99	81	46	214	−22	73	72
105	64	20	25	160	10	−49	52	215	− 4	− 7	49
106	−96	85	39	161	63	−41	39	216	54	−48	58
107	93	−29	42	162	38	39	94	217	−70	39	84
108	−40	−84	77	163	−28	39	97	218	54	−82	58
109	86	35	68	164	− 2	−47	18	219	29	41	41
110	91	36	50	165	38	8	3	220	−87	51	98
111	62	− 8	42	166	−42	− 6	23	221	−96	−36	77
112	−24	4	71	167	−67	88	19	222	49	8	57
113	11	96	85	168	19	93	40	223	− 5	54	39
114	−53	62	78	169	40	27	49	224	−26	43	99
115	−28	−71	64	170	−61	56	96	225	−11	60	83
116	7	− 4	5	171	43	33	58	226	40	61	54
117	95	− 9	93	172	−18	−39	15	227	82	35	86
118	− 3	17	18	173	−69	19	21	228	−92	12	2
119	53	−90	38	174	75	−18	56	229	−93	−86	14
120	58	−19	29	175	31	85	67	230	−66	63	42
121	−83	84	81	176	25	58	10	231	−72	−87	14
122	− 1	49	4	177	−16	36	36	232	−57	−84	55
123	− 4	17	23	178	91	15	84	233	23	52	2
124	−82	− 3	11	179	60	−39	59	234	−56	−62	18
125	−43	47	86	180	49	−47	85	235	−19	59	17
126	6	− 6	2	181	42	33	60	236	63	−14	22
127	70	99	31	182	16	−81	33	237	−13	38	28
128	68	−29	54	183	−78	53	62	238	−19	87	3
129	−94	−30	87	184	53	−80	70	239	44	−84	96
130	−94	−20	17	185	−46	−26	79	240	98	−17	53
131	−21	77	81	186	−25	−54	98	241	−16	62	15
132	64	37	72	187	69	−46	99	242	3	66	36
133	−70	−19	10	188	0	−78	18	243	26	22	98
134	88	65	50	189	−84	74	55	244	−38	−81	78
135	2	29	25	190	−16	16	75	245	70	80	92
136	33	57	71	191	−63	−14	94	246	17	−35	65
137	−70	6	85	192	51	−77	89	247	96	−83	64
138	−38	−56	51	193	−39	61	13	248	−77	80	43
139	−80	−95	29	194	5	97	19	249	−14	44	50

Vol. 4
1976

TABLE 4. MULTI-TERMINAL SWEEP ALGORITHM SOLUTION OF PROBLEM ELEVEN OF TABLE 1

Terminal	Route No.	Route	Distance	Load
1	1	92–178–210–41–110–60–158–109	114·2307	472
1	2	20–227–29–97–132–59–171–105	140·6073	485
1	3	85–169–181–162–75–52–243–16–6–165	145·9844	463
1	4	222–37–116–126–215–74–104–141–25–199–44–111	179·0652	489
1	5	196–174–143–14–107–128–48–23–120–22–236	121·2726	459
1	6	98–240–117	68·91753	230
2	1	17–24–40–127–147–159–134–51–67–32	172·2452	449
2	2	176–233–39–102–15–219–136–226–78–201	131·6328	499
2	3	242–225–235–82–157–43–223–122–38	146·1150	496
2	4	197–42–10–238–131–214–57–203–11–241–56	166·0002	488
2	5	175–66–80–194–113–103–168–47–55	84·93602	357
3	1	245–88–247–149–86–151–94–119	136·0695	493
3	2	192–184–218–50–77–212–45–91–144	141·7722	435
3	3	216–46–93–187–161–179–200–63–180–152	124·4228	491
3	4	156–28–95–188–87–34–182–35–239	128·5520	439
4	1	69–177–237–249–62–146–84–135–3–118–123–190	163·5731	486
4	2	163–224–64–125–193–101–114–195	131·1170	485
4	3	2–170–148–26–13–167–89–153–9–230–27	157·7891	472
4	4	183–58–248–121–106–189–150–36	165·1048	499
4	5	217–83–220–206–65–53–228–137–173	148·6997	490
4	6	204–198–73–96–112–8–213–166–207–30–18–54	141·8514	500
5	1	19–185–140–7–246–208–160–164–172–72	187·5112	450
5	2	186–81–4–115–145–100–21	136·4537	471
5	3	138–209–244–108–154–76–155–70–99–234–33	186·4304	496
5	4	79–68–142–232–231–202–139–1–229–71–31–205	207·8954	487
5	5	211–90–61–221–129–130–124–49–12–133–5	112·4839	431
5	6	191	13·41641	94
		Total 3754·147		

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

A Model with Shortage of Places for Educational and Manpower Systems

J MOYA-ANGELER[1]

(*Received September 1975; in revised form June 1976*)

If in a system the size of some grades either has or is going to outgrow some desired value, it will be necessary to impose some restraints, in the form of capacity limits, to the size of the grades. In this paper a stochastic model in discrete time for such kind of educational and manpower systems is developed. Here, the flow of people to some grades will not be determined by the number of people able to be promoted or recruited, but by the number of vacancies in such grades. A set of equations is derived to obtain the equilibrium structure of such systems. A new technique is explained for treating steady-state control through recruitment in a Markov system.

Vol. 4
1976

1. INTRODUCTION

TWO MAIN kinds of models have been developed in the field of educational and manpower systems, Markov models and Renewal processes. In both of them the elements of the system are grouped in different grades according to some characteristics. Transitions take place between grades according to some probabilistic laws. In the general case movements also occur between the system and the outside world, in the form of recruitment in the system and wastage from it. In the Markov-models the sizes of the grades are random variables while the transition probabilities are fixed. This kind of model is mainly concerned with the expected size of the grades at different times. Models of this kind have been developed by Young and Almond [6], Gani [5], etc., and Bartholomew [3] has studied the statistical properties of the general model. In the second kind of model the size of the grades is fixed and the rates of promotion are the random variables. Models of this kind have been treated by Bartholomew [2, 3], and Butler [4] among others (for further details of both kinds of models see Bartholomew [3]).

In practice, many systems behave in an intermediate way between these two extremes. For instance, in an educational system, the number of primary students is not restricted while the number of places for university students can be, and

[1]London School of Economics and Political Science.

in fact is, restricted. Several models have been developed to deal with these intermediate systems. Young and Vassiliou [7] have developed a model in which the promotions are not only determined by the number of people available to be promoted but also by the number of vacancies to be filled. Armitage, Smith and Alper [1] have gone a step further by defining what they called a Bottleneck system. In this kind of model it is assumed that in at least one grade "The flow is not simply the result of the demand for places but *may* be determined by the supply of places".

In this paper a third type of model is developed. Here we shall assume that from a certain time onwards the promotions to some grades will be determined only by the number of vacancies. This is because the size of these grades is not free but limited to a maximum size; i.e. these grades have capacity limits. When the demand for places in a grade is greater than the capacity limit we shall say that there is an overspill[2] in that grade. The overspill will be the excess of the demand over the supply. This kind of system can be considered as a special case of a Bottleneck system. In fact, what we shall do will be to strengthen the hypothesis about promotions by assuming that when a grade has a capacity limit this value must be reached and from then on maintained. This implies that from a certain time there will be shortage of places in the system. If when we place a capacity limit on a grade it is either because the size of the grade has grown in a way that can disturb the behaviour of the system or because we can foresee that the size of the grade is going to outgrow some desired value, the hypothesis made will be a reasonable one.

As is usual in manpower systems we shall assume that promotions can only occur one step ahead and that demotions never occur (the theory also applies when we allow for transitions between all grades, but it must then be clearly stated how the overspills are going to be reallocated, since for each policy of reallocation the structure of the lower grades will be modified). The system is assumed to behave as a Markov model until the capacity limit is reached in some grade. From then on the size of the grade will be its capacity limit. The number of promotions and recruits to this grade will be determined by the number of vacancies arising in such a grade. From the total number of people able to be promoted there will therefore be a number that cannot be promoted. Of this number of frustrated promotions there will be a proportion that decide to stay in the system, and remain in the grade in which they were. The rest will leave the system and will increase the wastage of the grade from which they came. The process will continue in the same way until all the capacity limits have been reached. From that point on, the size of some grades will be random variables and the others will be fixed. Section 4 deals with the equilibrium structure of the expectations of these random variables. Section 6 is concerned with the conditions that the sizes of some grades must satisfy in order for the promotions and

[2] Notice that supply and demand refer, in this paper, to vacancies rather than people.

recruits to be determined by the supply of places. All these results are applied in Section 7 to the steady-state control through recruitment on a Markov-model.

2. NOTATION

We shall consider a system with k grades supposing that the grades are arranged in increasing order of seniority.

$\bar{n}_i(T)$ denotes the expected number of elements in grade i at the end of the interval $(T-1, T)$ (in short, time T). The same symbol without T will denote the value at equilibrium;

$N(T)$ denotes the total number of elements in the system at the end of the interval $(T-1, T)$;

P_{ij} denotes the probability that a member of grade i moves to grade j, assuming that there is no capacity limit at grade j;

$R_i(T)$ denotes the number of new entrants to grade i at the end of the interval $(T-1, T)$;

w_i denotes the probability that an element of grade i leaves the system;

μ_i denotes the capacity limit of grade i;

β_i denotes the proportion of frustrated promotions that decide to remain in the system.

Vol. 4
1976

3. BASIC RELATIONS

The expected structure of the system at time T can be obtained from the following set of equations:

$$\begin{aligned}\bar{n}_i(T) &= [\text{stayers from } T-1] + [\text{inflow at } T] \\ &= (P_{ii} + \beta_i\delta_{i+1}(T))\bar{n}_i(T-1) + \alpha_i(T)[P_{i-1,i}\bar{n}_{i-1}(T-1) + R_i(T)] \\ &\qquad i = 1, \ldots, k \end{aligned} \tag{3.1}$$

where $\delta_{i+1}(T)$ denotes the proportion of people able to be promoted from grade i but cannot be allocated in grade $i+1$ at time T. It is defined by:

$$\delta_{i+1}(T) = (1-\alpha_{i+1}(T))P_{i,\,i+1}. \tag{3.2}$$

When the grade $i+1$ has capacity limit, $\alpha_{i+1}(T)$ is the ratio between supply and demand, i.e.

$$\alpha_{i+1}(T) = \frac{[\text{Supply}]}{[\text{demand}]} = \frac{[\text{Vacancies}]}{[\text{potential inflow}]} = \frac{\partial_{i+1}(T)}{P_{i,\,i+1}\bar{n}_i(T) + R_{i+1}(T)}. \tag{3.3}$$

For simplicity it is assumed that the recruits and the promotions have the same chance of getting a place. When the grade $i+1$ has no capacity limit $\alpha_{i+1}(T) = 1$.

The number of vacancies $\partial_i(T)$ in grade i at time T is given by:

$$\begin{aligned}\delta_i(T) &= [\text{capacity limit}]-[\text{stayers from } T-1] \\ &= \mu_i-(p_{ii}+\beta_i\delta_{i+1}(T))\bar{n}_i(T-1)\,. \end{aligned} \tag{3.4}$$

The wastage of the grade, immediately before one with a capacity limit, will be increased by the number of frustrated promotions that decide to leave the system. The total wastage of grade i at time T will thus be

$$\begin{aligned}\Omega_i(T) &= [\text{usual leavers}]+(1-\beta_i)[\text{overspill promotees}] \\ &= w_i\bar{n}_i(T)+(1-\beta_i)\theta_{i+1}(T)\end{aligned} \tag{3.5}$$

where the overspill at grade $i+1$ is defined by

$$\theta_{i+1}(T)=\delta_{i+1}(T)\bar{n}_i(T)\,. \tag{3.6}$$

The set of equations defined by (3.1) must be solved recursively at each time T, beginning at the top grade in order to know whether an overspill is produced.

All these relationships are graphically represented in Fig. 1.

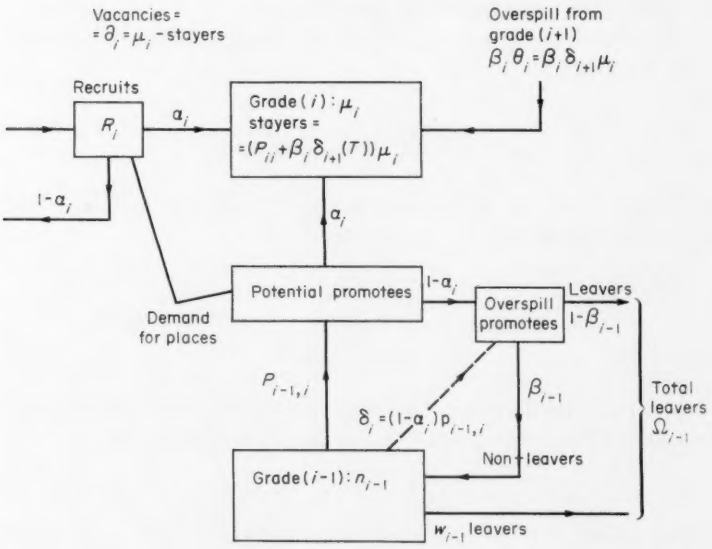


FIG. 1.

4. STEADY-STATE

The method developed by Bartholomew [3] to obtain the steady-state of a Markov-system is not applicable here because at time $T-1$ we do not know what

the overspills at time T are going to be. The equilibrium structure of a system with k grades, in which it is assumed that there are capacity limits from the grade $r + 1$ up to the grade $r + s$ but not in the others, is derived in the Appendix. There is constant recruitment in all the grades. Any system with capacity limits can be obtained as a particular case of this one.

When the capacity limits have been reached, the grades of the system can be divided into the following three groups:

1. Grades with capacity limits, whose sizes will be determined by their limits.
2. Grades without capacity limits, but directly influenced by these limits. These grades are: the one immediately before the first one with a capacity limit in which there will be an overspill, and the grade following the last one with a capacity limit in which the number of elements promoted to it will be constant, from the time that the capacity limit has been reached in the grade below.
3. Grades without capacity limits and outside the direct influence of the limits, whose behaviour will be the same as in a Markov model.

The steady-state of the system can be obtained using the following three facts:

(i) Independence property: A grade (or set of grades) with capacity limits divides the system in two[3] subsystems without interaction.

(ii) Bartholomew [3] has proved that the equilibrium structure when recruitment is constant always exists. What is done in this paper is to approach (as much as desired) the equilibrium structure.

(iii) The equilibrium is reached in a recursive way beginning at the bottom grade.

Vol. 4
1976

5. EXAMPLE

To illustrate this theory we shall give a simple example. A program has been written that computes, using (3.1), the successive structures and the steady state of any system. The program has been applied to an imaginary firm with 4 grades. It is assumed that there are capacity limits in the third and fourth grades. These limits are 950 and 350 respectively. The proportion of frustrated promotions (β_i) that remain in the system is the same for all the grades and equal to $\frac{1}{2}$. The total recruitment to the system at each time is 1000 distributed in the following way (650, 200, 100, 50). The initial structure of the system is (2400, 1600, 900, 300). Movements take place in the system according to the following matrix.

$$P = \begin{bmatrix} 0{\cdot}6 & 0{\cdot}25 & 0 & 0 \\ 0 & 0{\cdot}75 & 0{\cdot}13 & 0 \\ 0 & 0 & 0{\cdot}83 & 0{\cdot}08 \\ 0 & 0 & 0 & 0{\cdot}95 \end{bmatrix}.$$

[3] The set of equations with capacity limits is not taken into account here.

The steady-state is supposed to be reached when the difference between two successive values is less than 0·01 for all grades. The structure of the system at times $T = 1$, $T = 24$, and $T = 47$ are printed in Table 1. The reason for printing these values is that at these times the limits and the equilibrium values, to the accuracy specified, are reached. The equilibrium structure which was obtained by using the theory is also printed.

TABLE 1. EXPECTED GRADE SIZES

Grades	Capacity limits	$T=0$	$T=1$	$T=24$	$T=47$	Theoretical steady-state
1	—	2400	2090 (1)	1625 (1)	1625 (1)	1625
2	—	1600	1881 (1)	2995 (1)	3000 (1)	—
3	950	900	950 (0·66)	950 (0·263)	950 (0·263)	950
4	350	300	350 (0·53)	350 (0·14)	350 (0·14)	350

The numbers in brackets are the proportions $a_i(T)$.

6. CONDITIONS FOR THE FLOW TO BE DETERMINED BY THE NUMBER OF VACANCIES

We have assumed that once the capacity limits have been reached the flow of people to these grades will be determined by the supply of places. That means that the demand for places will be greater than the supply and therefore there will be an overspill in such grades. The demand for places in a grade depends on the number of people able to be promoted and on the number of elements that want to enter the system in this grade. As we have supposed that the recruitment is constant, the demand for places will thus depend on the number of people able to be promoted. As the probabilities of promotion are constant the demand will only depend on the size of the previous grades. We are going to see what conditions the sizes of the grades must satisfy so that the demand will always be greater than the supply.

The set of grades with which these conditions are concerned is the one in which the overspills are reallocated. The grades $r + 2, \ldots, r + s$ will be at their capacity limits when:

$$[\text{demand} \geq \text{supply}] \Leftrightarrow [\text{potential inflow}]_i \geq [\text{vacancies}]_i$$

$$i = r + 2, \ldots, r + s - 1$$

$$\Leftrightarrow [p_{i-1,i}\mu_{i-1} + R_i] \geq \mu_i[1 - p_{ii} - \beta_i\delta_{i+1}]$$

$$\Leftrightarrow \mu_{i-1} \geq \frac{\mu_i(1 - p_{ii} - \beta_i\delta_{i+1}) - R_i}{p_{i-1,i}} \,. \tag{6.1}$$

For the first grade with overspill we can obtain a similar equation to (6.1).

$$\bar{n}_r(T) \geq \frac{\mu_{r+1}(1-p_{r+1,\,r+1}-\beta_{r+1}\delta_{r+2})-R_{r+1}}{p_{r,\,r+1}} \,. \tag{6.2}$$

The problem now is that as the size of the r'^{th} grade is not constant we do not know if the possible values of $n_r(T)$ are going to satisfy (6.2). The fact that the RHS of (6.2) is constant provides us with a bound for the lowest value that the size of the grade r can take, in order that the size of the grade $r+1$ will be the capacity limit. When the initial value is greater than the bound, there is a way in which we can know *a priori* if the overspill is going to be maintained. First we compute the set of equations (6.1) to determine whether the overspills are maintained. If this is the case we compute the equilibrium structure by assuming that the size of the grade $r+1$ is the capacity limit. If $\bar{n}_r$ is greater than the bound the overspill in this grade will be maintained.

Again, these results can be applied to obtain the conditions for any system with any capacity limit distribution. When there is only recruitment at the bottom the relations are:

$$\left.\begin{aligned} &\mu_{i-1} \geq \mu_i \;\; \frac{1-p_{ii}-\beta_i\delta_{i+1}}{p_{i-1,\,i}}; \;\; i = r+2, \ldots, r+s+1 \\ &\bar{n}_r(T) \geq \mu_{r+1} \;\; \frac{1-p_{r+1,\,r+1}-\beta_{r+1}\delta_{r+2}}{p_{r,\,r+1}} \end{aligned}\right\}. \tag{6.3}$$

We must point out that if (6.2) does not hold the size of the grade $r+1$ will be less than its capacity limit and this can influence the upper grades. To illustrate this fact a simple example can be given.

The data are the same as in Section 5, except that the recruitment is only at the bottom grade with a value of 400 and the capacity limits are now in the second and third grades with values of 1600 and 950 respectively. From (6.2)[4] we obtain that $n_1(T) \geq 1507$ to maintain the capacity limit in the second grade. By the program we obtain that $n_1(2) = 1559$. As this number is greater than the bound there is still an overspill at time 3.[5] The size of the second grade is the capacity limit. At this time $n_1(3) = 1343$ which implies that when $T = 4$ there is no overspill in the second grade and μ_2 is no longer maintained. Indeed, $n_2(4) = 1559$ and the capacity limit was 1600. The first equation of (6.3) would hold if μ_2 is maintained, because $1600 \geq 1242$. But as μ_2 is not maintained this influences the size of the third grade. In fact, when $T = 8$ $n_2(8) = 1273$, therefore when $T = 9$ there is still an overspill in the third grade. At this time $n_2(9) = 1214$, so that when $T = 10$ the size of the third grade is less than the capacity limit. These values are 946 and 950 respectively.

[4] As in this case the first grade with capacity limit is the second one, apply (6.2) instead of the r'^{th} equation of (6.3).

[5] We must state that in (6.2) and in the r'^{th} equation of (6.3) the value of μ_{r+1} is at time $T+1$. This is not printed because μ_{r+1} is constant.

Vol. 4
1976

7. APPLICATION TO CONTROLLING THE STRUCTURE OF A GENERAL MARKOV MODEL

Control theory for Markov models has been developed in order to know, once a desired structure has been decided upon, whether or not this structure can be attained and maintained. As we want to control the number of elements in each grade the method of control is the flow of people through the system. Of the three flows (wastage, promotion and recruitment), control through recruitment seems to be less problematic because decisions taken upon it do not affect the people already in the firm (see Bartholomew [3]). The control can be exercised in two different ways either by a fixed set of control parameters or by a sequence of such parameters. The first way is known as steady-state control and the second one as sequential control. In the case of steady-state control the target structure must be attained as the equilibrium structure of the system otherwise the structure will not be maintained (in the rest of the paper we shall refer to this type of steady-state control as control without restraints). The fact that the equilibrium structure of a Markov model with upper triangular transition matrix tends to be top-heavy (i.e. a great part of the elements of the system are accumulated in the highest grade) will imply that the structures that can be maintained will be the top-heavy ones. Bartholomew [3] has proved this fact by obtaining the maintainable region for this type of control.

In this section a new type of steady-state control is developed. The idea is based on the fact that in a renewal process a total control is exercised over the system, and that the model developed here is an intermediate one between a Markov model and a renewal process. We will see that by choosing in a convenient way the grades on which we are going to place the capacity limits, and exercising control through recruitment, any desired structure that satisfies the conditions expressed in Section 6 can be attained and maintained. We shall take as capacity limits in all the grades except the first one,[6] the desired values of the grades. The desired size of the first grade is attained as the equilibrium value of the grade. We will prove that for this grade it is possible to attain any value. We can control the system either by making all recruits enter at the lowest grade or by allowing lateral recruitment.

We know that the limiting value of the first grade when there is lateral recruitment is given by:

$$\bar{n}_1^2 p_{12} D + \bar{n}_1[R_2 D + \beta_1 \partial_2 p_{12} - p_{12} R_1] - R_1 R_2 = 0$$

where we have put $D = 1 - p_{11} - \beta_1 p_{12}$. Therefore:

$$R_1 = \frac{\bar{n}_1^2 p_{12} D + \bar{n}_1[R_2 D + \beta_1 \partial_2 p_{12}]}{p_{12}\bar{n}_1 + R_2} . \tag{7.1}$$

[6] It can be proven that any system with any capacity limit distribution can be controlled in this way.

Vol.
197

It is straightforward to prove that all the terms on the RHS are positive. Therefore, there exists a positive value of R_1 that produces $\bar{n}_1$ as the limiting value.

From (7.1) the value of R_1 can be obtained as a particular case when there is only recruitment at the bottom

$$R_1 = \bar{n}_1 D + \beta_1 \partial_2$$

and again the same occurs. As the desired structures are obtained as equilibrium structures it is obvious that they will always be maintained.

The main advantage of this technique is that not only does it allow us to control the top-heavy structures, but it also makes possible the control of the non-top heavy structures. Indeed, from (6.1) and (6.2) it is clear that any system in which the size of the grades decreases as we ascend in the hierarchy can be controlled in this way. This method therefore involves the control without restraints, allowing us to attain and maintain a greater range of structures. We are going to show with an easy example how a structure that is attainable and maintainable by using this technique cannot be maintained by using the steady-state control proposed in Bartholomew [3].

The target structure is (2000, 1500, 1000, 500) or in relative sizes (0·4, 0·3, 0·2, 0·1). The initial state is (1600, 1200, 700, 400). The transition matrix is the same as in Section 5. In the control without restraints a structure is maintainable if the vector **r** obtained by solving

$$\mathbf{r} = \mathbf{q}(\mathbf{I} - \mathbf{P})/\mathbf{q}\mathbf{w}' \tag{7.2}$$

has all its elements positive. **q** is the vector of relative size of the grades and **r** is the vector of the distribution of the recruitment. If we solve (7.2) in this case the vector obtained is

$$\mathbf{r} = (1{\cdot}3445, -0{\cdot}210084, -0{\cdot}04201, -0{\cdot}0924369).$$

So, the structure cannot be maintained.

On the other hand, if we apply the results obtained above, one of the possible recruitment vectors to maintain the target structure is

$$\mathbf{R} = (665, 200, 100, 50).$$

The system has been simulated by the program and with this vector of recruitment the capacity limits have been reached in the second year (all at the same time). The target structure has been reached (in the sense defined in Section 4) in the 30th year. The figures obtained are shown in Table 2.

8. CONCLUSIONS

This paper provides a simple way of knowing what the behaviour of a system with shortage of places is going to be like, when the policies exercised upon the

Vol. 4
1976

TABLE 2. STEADY-STATE CONTROL WITH RESTRAINTS

Grades	Capacity limits	$T = 0$	$T = 2$	$T = 30$	Target
1	—	1600	1718 (1)	1998 (1)	2000
2	1500	1200	1500 (0·61)	1500 (0·46)	1500
3	1000	700	1000 (0·96)	1000 (0·47)	1000
4	500	400	500 (0·32)	500 (0·19)	500

system are maintained over a period of time. We believe a more comprehensive technique of treating the steady-state control of a general system has been developed. The method involves both an explicit recruitment control and an implicit promotion control allowing us to attain a greater number of structures. It would be interesting to generalise the model in order to allow expansion or contraction in the capacity limits. This paper is only a first step in the theoretical study of bottleneck systems and as such there is still much work to be done in the study of more general cases.

Vol. 197

ACKNOWLEDGEMENTS

The author gratefully acknowledges an immense debt to Professor David J. Bartholomew for much help and encouragement. This work has been undertaken with the financial support of the 'Fundacion Juan March'.

REFERENCES

1. ARMITAGE PH, SMITH CS & ALPER P (1969) *Decisions Models for Educational Planning.* Penguin Press, London.
2. BARTHOLOMEW DJ (1969) Renewal theory models for manpower systems. In *Manpower Research* (Ed. WILSON NAB) pp. 120–128. English University Press, London.
3. BARTHOLOMEW DJ (1973) *Stochastic Models for Social Processes.* Wiley, New York.
4. BUTLER AD (1970) Renewal theory applied to manpower systems. Ph.D. thesis, University of Kent.
5. GANI J (1963) Formulae for projecting enrollments and degrees awarded in universities. *J. R. Statist. Soc.* **126,** 400–409.
6. YOUNG A and ALMOND G (1961) Predicting distributions of staff. *Computer J.* **3,** 246–250.
7. YOUNG A and VASSILIOU P-CG (1974). A non-linear model on the promotion of staff. *J. R. Statist. Soc.*A. **137,** 584–595.

ADDRESS FOR CORRESPONDENCE: *J. Moya-Angeler, D-6 Westgate, Cambridge, MA 02139, USA*

APPENDIX

MATHEMATICAL DERIVATION OF THE STEADY-STATE

According to the classification of the grades stated in Section 4, the equations of the system can be written down in the following way:

$$\left.\begin{aligned}
&\bar{n}_1(T) = p_{11}\bar{n}_1(T-1) + R_1 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\bar{n}_i(T) = p_{ii}\bar{n}_i(T-1) + p_{i-1,\,i}\bar{n}_{i-1}(T-1) + R_i,\ i = 2, 3, \ldots, r-1 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\bar{n}_r(T) = p_{rr}\bar{n}_r(T-1) + \beta_r\delta_{r+1}(T)\bar{n}_r(T-1) + p_{r-1,\,r}\bar{n}_{r-1}(T-1) + R_r \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}\right\} \qquad (A.1)$$

$$\left.\begin{aligned}
&\bar{n}_j(T) = \mu_j;\ j = r+1, \ldots, r+s \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}\right\} \qquad (A.2)$$

$$\left.\begin{aligned}
&\bar{n}_{r+s+1}(T) = p_{r+s+1,\,r+s+1}\bar{n}_{r+s+1}(T-1) + p_{r+s,\,r+s+1}\mu_{r+s} + R_{r+s} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\bar{n}_l(T) = p_{ll}\bar{n}(T-1) + p_{l-1,\,l}\bar{n}_{l-1}(T-1) + R_l;\ l = r+s+2, \ldots, k
\end{aligned}\right\} \qquad (A.3)$$

Notice that in the equation that defines the grade $r+s+1$ there is no influence from the first r grades. As the number of promotions to this grade is constant, we can treat this grade as the first one of a system in which the recruitment to this grade has been increased by the number of promotions.

We can approach the steady-state as near as we want in a finite time. For each grade i we can choose an ϵ, as small as we want, such that the difference between the size of two consecutive values will be less than ϵ, and when this happens we will say that the grade has reached the steady-state.

Grade i will not be able to reach equilibrium unless equilibrium has been reached in grade $i-1$. If this was not so the number of promotions from grade $i-1$ to grade i would not be constant. This is also true for grade r, because in the overspill the only variable is $\bar{n}_r(T)$. Therefore, the equilibrium is reached (or approached) in a recursive way beginning at the bottom grade.

The equation

$$\bar{n}_1(T) = p_{11}\bar{n}_1(T-1) + R_1$$

can be seen as an iterative process to solve the equation

$$\bar{n}_1 = p_{11}\bar{n}_1 + R_1 .$$

But obviously this equation has an exact solution

$$\bar{n}_1 = \frac{R_1}{1-p_{11}} . \qquad (A.4)$$

One the equilibrium has been reached in grade $i-1$, the size of the grade i at equilibrium will be given by

$$\bar{n}_i = \frac{p_{i-1,i}\,\bar{n}_{i-1} + R_i}{1-p_{ii}}\ ;\ i = 2, 3, \ldots, r-1 . \qquad (A.5)$$

By using (3.2), (3.3) and the r'^{th} equation of (A.1) and bearing in mind that the number of vacancies in grade $r+1$ is constant, the size of grade r in equilibrium is given by one of the

Vol. 4
1976

solutions of the quadratic equation:

$$\begin{aligned}&\bar{n}_r^2 p_{r,\,r+1}(1-p_{rr}-\beta_r p_{r,\,r+1}) + \bar{n}_r[R_{r+1}(1-p_{rr}-\beta_r p_{r,\,r+1})\\&\quad+\beta_r \partial_{r+1} p_{r,\,r+1} - \bar{n}_{r-1} p_{r-1,\,r} p_{r,\,r+1} - p_{r,\,r+1} R_r]\\&\quad-R_{r+1}[p_{r-1}\bar{n}_{r-1} + R_r] = 0\,.\end{aligned} \tag{A.6}$$

It is quite easy to prove that the roots of this equation have different signs, and therefore there is no indeterminacy. The number of vacancies in the grade $r+1$ can be obtained by successive application of (3.2), (3.3), and (3.4). It is given by

$$\partial_{r+1} = \Gamma_{r+1} - \beta_{r+1}\Delta_{r+2}\left[1 - \frac{\partial_{r+2}}{\rho_{r+2}}\right] \tag{A.7}$$

this equation must be solved recursively up to the last grade with capacity limit in which the number of vacancies is

$$\partial_{r+s} = \Gamma_{r+s}\,.$$

Where

$\Gamma_l = \mu_l(1-p_{ll})$ = ['potential' movers];

$\Delta_l = p_{l-1,\,l}\,\mu_{l-1}$ = ['potential' promotees];

$\rho_l = \Delta_l + R_l$ = ['potential' inflow].

The equilibrium structure of (A.3) can be obtained in a similar way; it is given by

$$\left.\begin{aligned}\bar{n}_{r+s+1} &= \frac{p_{r+s,\,r+s+1}\mu_{r+s} + R_{r+s+1}}{1-p_{r+s+1,\,r+s+1}}\\ \bar{n}_l &= \frac{p_{l-1,\,l}\bar{n}_{l-1} + R_l}{1-p_{ll}}\quad ; l = r+s+2, \ldots, k\end{aligned}\right\}. \tag{A.8}$$

A very common situation is when all the new elements enter the system at the bottom grade, in this case a simpler set of equations can be obtained.

From (A.5), (A.6) and (A.8) the equilibrium structure of any system with any distribution of the capacity limits and with lateral recruitment can be obtained. Each one of these equations will be used according to the position of the grade in the system; i.e. the equilibrium structure of a grade without capacity limit and immediately before one with a capacity limit will be given by (A.6) For the grade following the last one with a capacity limit the first equation of (A.8) must be applied. The steady-state of the grades with capacity limits will be given by their limits. For the rest of the grades (A.5) can be applied. If one grade without capacity limit is situated between two with capacity limit, (A.6) can be applied by changing $\bar{n}_{r-1}$ and μ_{r-1}.

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Memoranda

A New Goal Programming Formulation

THE STANDARD formulation of a goal programme with non-preemptive weights is [1]:

$$\left.\begin{aligned} \text{Min. } Z &= \sum_i w_i^+ u_i^+ + w_i^- u_i^- \\ \text{s.t. } \mathbf{r}_i\mathbf{x} - u_i^+ + u_i^- &= g_i \;\forall\, i \\ \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{x},\ \mathbf{u}^+,\ \mathbf{u}^- &\geq \mathbf{0} \end{aligned}\right\} \quad (1)$$

where $\mathbf{w}^+$ and $\mathbf{w}^-$ are the vectors of non-preemptive weights on the deviation above and the deviation below the goal vector $\mathbf{g}$ respectively and $\mathbf{x}$ is the vector of decision variables. The formulation implies that the user of the goal programme is indifferent between deviations of $u_j = 1/w_j$ and $u_k = 1/w_k$ where $k \neq j$ and the overage and underage superscripts have been omitted as being irrelevant. (Notice also that it is implicitly assumed that the weights are positive although the subsequent discussion may be easily extended to the situation where there are zero or negative weights.)

To find a basic feasible solution to this problem is straightforward;

1. Find a basic feasible solution, $\mathbf{x}_B$, to $\mathbf{A}\mathbf{x} = \mathbf{b}$ by a Phase 1 of the Simplex algorithm;

2. Calculate $e_i = g_i - \mathbf{r}_i\mathbf{x}_B \;\forall i$ and set $u_i^+ = -e_i$ if $e_i < 0$, $u_i^- = e_i$ if $e_i > 0$, whilst keeping all the remaining u's zero.

Thus a basic feasible solution is $(\mathbf{x}_B, \mathbf{u}_B)$, where $\mathbf{u}_B$ is the positive vector of basic deviations.

The above formulation has been widely described in the literature, e.g. [2], and a very large number of theoretical papers, and a somewhat smaller number relating to real applications, have been published. This note outlines what happened when the author presented results based upon this formulation to a particular company; they were rejected on two inter-connected grounds. The first cause for complaint was that, because the above model is in essence a linear programme. the optimal solution for a particular set of weights lay at one vertex of the feasible space; if the set of weights were changed (only very slightly in some cases), the optimal solution moved to a different vertex, The practical implication of this was that although the numerical values of the weights could be varied continuously, the dissatisfaction or otherwise of the goals would only change in discrete amounts (assuming that alternate solutions do not exist). This situation was not acceptable to the management; they felt that if they, say, increased the value of one weight in isolation, then there should be an immediate reduction in the dissatisfaction of the goal associated with that weight (assuming of course that the goal was originally dissatisfied) relative to the dissatisfactions of the other goals. The second related cause was that, again as a direct result of the Simplex algorithm, generally those deviations with the highest associated weights were set to zero in the optimal solution, whilst those deviations with the lowest weights had relatively high values. The management thought that, in specifying the weights, they were quantifying the relative trade-offs between *deviations* from goals, not between satisfaction or dissatisfaction of goals. As an example, consider the table below:

	Goal		Z
	1	2	—
Weights	2	1	—
Case 1: $\mathbf{u}$	0	1	1
Case 2: $\mathbf{u}$	1/3	2/3	4/3

Vol. 4
1976

Case 1 could be the solution produced by the goal programming formulation; it would however be unacceptable because the deviations bear no direct relationship with the relative trade-offs whereas Case 2, albeit resulting in a larger value of Z, would have been much more acceptable to the company.

As a result of these discussions, the following formulation was produced:

$$\text{Min. } Z = \qquad \gamma$$

$$\text{s.t. } \mathbf{r}_i\mathbf{x} - u_i^+ + u_i^- = g_i$$

$$\mathbf{A}\,\mathbf{x} = \mathbf{b}$$

$$w_i^+ u_i^+ + w_i^- u_i^- - \gamma + s_i = 0$$

$$\mathbf{x}, \quad \mathbf{u}^+, \quad \mathbf{u}^-, \quad \gamma, \quad \mathbf{s} \geq \mathbf{0}$$

where γ is the maximum weighted deviation from any particular goal, and $\mathbf{s}$ is a vector of slack variables measuring the difference between this maximum and the other deviations.

This new formulation essentially concentrates the algorithm's attention on those dissatisfactions that are greater than would be expected from the relative values of the weights and attempts to bring them back into line. In doing so, those dissatisfactions that are less than expected may be increased *pari passu*. The formulation is equivalent to $Z^* = \min_{x} \max_{i} (w_i^+ u^+ + w_i^- n_i^-)$ subject to (1). If we consider the criticisms of the original formulation expressed above, then the new model firstly possesses a continuous relationship between the deviation from a goal and the weight associated with it, and secondly attempts to maintain the relative trade-offs between the deviations. Certainly the model was now more acceptable although it is recognised that it may well possess different inherent weaknesses.

A basic feasible solution to this formulation may be found by adding, after the first two steps, the following further step:

3. Let $\mathbf{w}_B$ be the vector of weights corresponding to the elements of $\mathbf{u}_B$. Set $\gamma = \max_i (w_{B_i} u_{Bi})$. Suppose that this maximum corresponds to the kth element; then set $s_i = \gamma - w_{B_i} u_{Bi}$ (for $i \neq k$).

The new formulation described was a result of discussions with an individual company. This company was perhaps unusual in that it was mathematically orientated, that it had only a very few well-defined goals to be considered and that it had some very firm ideas as to the nature of the solution it wanted, but other companies may find the above discourse of some interest.

REFERENCES

1. KORNBLUTH JSH (1973) A survey of goal programming. *Omega* **1**(2), 193–205.
2. LEE SM (1972) *Goal Programming for Decision Analysis*. Auerbach, Philadelphia, PA, USA.

RB Flavell
(*June 1976*)
Department of Management Science
Imperial College of Science and Technology
Exhibition Road
London SW7 2BX
UK

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

A Graphical Interpretation in Goal Programming Problem

IN LEE [1], a technique developed for obtaining a graphical solution to the goal programming problem is discussed and illustrated. The step-by-step procedure to be followed in using the graphical method to obtain the optimum point is as follows:

"1. Identify and develop the model constraints from the problem.
2. Create new constraints if the deviational variable required in order to attain a certain stated goal is not present in the constraints already identified.
3. Develop the objective function by assigning appropriate priority and weight factors to deviational variables.
4. Plot the straight lines for all constraints involving choice variables.
5. Plot the straight lines for additional constraints that involve only deviational variables by converting it into a constraint involving choice variables.
6. Identify the point or area that satisfies the first goal, second goal, and so forth until the last goal is achieved to the fullest possible extent.
7. Identify the optimum point and analyze it in terms of degrees of goal attainment." [1], p. 91.

Although proper adherence to these instructions will indeed identify the optimal solution to the goal programming problem, the techniques proposed by Lee in accomplishing step (6) are ill-advised.

As an example of the techniques, the following problem is solved using the graphical method ([1], p. 86):

$$\text{Min } Z = P_1 d_1^- + P_2 d_4^+ + 4P_3 d_2^- + 3P_3 d_3^- + 4P_4 d_3^+ + 3P_4 d_2^+$$

$$\begin{aligned}
\text{s.t. } & 2x_1 + 1\tfrac{1}{2}x_2 + d_1^- - d_1^+ = 180 \\
& x_1 + d_2^- - d_2^+ = 40 \\
& x_2 + d_3^- - d_3^+ = 40 \\
& x_1 + d_4^- - d_4^+ = 45 \\
& x_1, x_2, d_1^-, d_2^-, d_3^-, d_4^-, d_1^+, d_2^+, d_3^+, d_4^+ \geq 0
\end{aligned}$$

Figure 1 illustrates the remaining feasible region (shaded area) after having satisfied the first three goals. The next step is to satisfy the fourth goal "to the fullest possible extent."

Lee's advice is as follows ([1], p. 88):

"Our last goal is to minimize the sum of overtime for both lines. However, the

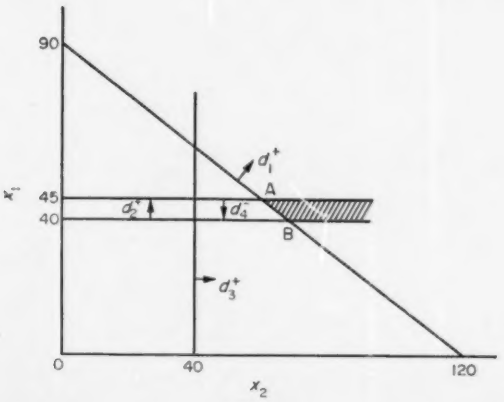


FIG. 1.

Vol. 4
1976

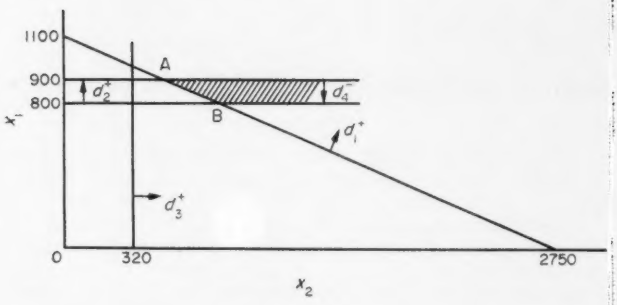


FIG. 2.

limitation of overtime for line 2 is given a greater weight. Therefore, the overtime operation of line 2 must first be minimized to the fullest extent. Obviously, at point A the overtime operation of line 2 will be minimized while meeting the first three goals. Next, we try to minimize the overtime operation for line 1. Since we cannot decrease the overtime operation for line 1 any further from point A without providing more overtime to line 2, point A has to be the optimum point."

Although this advice leads to the optimal solution in the examples presented by Lee, it is our contention that an optimal solution to the goal programming problem will not always be obtained using this procedure.

The following problem is a counter-example for this technique.

$$\text{Min } Z = P_1 d_1^- + P_2 d_4^+ + 2P_5 d_2^- + P_3 d_3^- + 2P_4 d_2^+ + P_4 d_3^+, \quad (1)$$

$$\text{s.t. } 5x_1 + 2x_2 + d_1^- - d_1^+ = 5500, \quad (2)$$

$$x_1 + d_2^- - d_2^+ = 800, \quad (3)$$

$$x_2 + d_3^- - d_3^+ = 320, \quad (4)$$

$$x_1 + d_4^- - d_4^+ = 900, \quad (5)$$

$$x_1, x_2, d_1^-, d_2^-, d_3^-, d_4^-, d_1, d_2^+, d_3^+, d_4^+ \geq 0. \quad (6)$$

According to Lee's technique, the optimum solution would be at point B in Fig. 2. This solution is $x_1 = 800$, $x_2 = 750$, $d_2^+ = 0$, $d_3^+ = 430$ and $Z = 430P_4$. But let us consider the problem after having just satisfied the third goal; the shaded area in Fig. 2 being the new feasible region. At this point the problem becomes more complicated. The object of the fourth goal is to drive the function $2d_2^+ + d_3^+$ as close to zero as possible while remaining in the feasible region indicated in Fig. 2. From inspecting the feasible region it is clear that d_2^+ and d_3^+ cannot be driven to zero simultaneously (in fact, no point in the feasible region corresponds to $d^+ = 0$).

From (3) and (4) it can be shown that

$$2d_2^+ + d_3^+ = 2x_1 + x_2 + 2d_2^- + d_3^- - 1920.$$

But d_2^- and d_3^- must both equal zero (as a result of the third goal). Hence the minimization of $2d_2^+ + d_3^+$ is equivalent to the minimization of $2x_1 + x_2$. The optimal point is point A with $x_1 = 900$, $x_2 = 500$, $d_2^+ = 100$, $d_3^+ = 180$ and $Z = 380P_4$.

Lee's approach attempts to minimize d_2^+ to the fullest extent before considering d_3^+ merely because the weight associated with d_2^+ is larger than the weight associated with d_3^+. This approach fails to recognize the tradeoff between d_2 and d_3^+ and instead treats the problem as if d_2^+ and d_3^+ have different preemptive priority factors.

It is suggested therefore that step (6) should be revised as follows: "Identify the point or area that satisfies the first goal, second goal, and so forth, until the last goal is achieved to the fullest possible extent. When two or more deviational variables are associated with the same preemptive priority

factor, express the function of the deviational variables which is to be minimized to the fullest extent in terms of the choice variables (x_1 and x_2 in the example given) by using the constraints given in the problem. Then, minimize the resulting function of the choice variables subject to the feasible region which has been created as a result of the process of satisfying higher priority goals to the fullest extent."

In summary, it is felt that the graphical procedure for goal programming aids in the comprehension of the purpose and intent of goal programming in that (1) it elucidates the meaning of minimizing the deviational variables; (2) illustrates that the hierarchical process of first satisfying higher level priorities serves as a set of constraints for subsequent priorities; and (3) demonstrates the nature of the tradeoff which exists between deviational variables within the same priority class.

REFERENCE

1. LEE SM (1972) *Goal Programming for Decision Analysis.* Auerbach, Philadelphia, PA.

EL Hannan
(*June 1976*)

Institute of Administration & Management
Union College & University
Schenectady, New York 12308
USA

Vol. 4
1976

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Using the Delta Chart in Project Management

FOR THE project director of a large research project, several critical decisions which he has to make involve the planning and monitoring of the research design. As the chief decision-making for the project, he must have a clear concept of what activities and events are necessary within a given length of time. In addition he must identify key decision points for data collection in the research design, allocate resources and modify the research strategy to fit prevailing circumstances.

Of critical importance to the project director is the need for effective communication between himself and the other persons working on the research study. Since the problem under attack is complex and the project lengthy, it is likely that the study will employ the expertise of many professionals—a team approach with coordinated activities. Although their efforts may be done individually, they are in the final analysis working together as members of the research team. The project director must decide which of these activities must come first, which must be finished before another can begin, and which ones may be carried out at the same time. In essence he needs an overview of the set of research activities to be carried out, and he needs to be able to illustrate the interdependency of the individually assigned project activities.

What is needed is a communication device that would clearly and concisely display the research project. Such a device would ideally visually display an overview of the research project that could be grasped in a few moments of study by each research participant; show the connecting interdependencies between each activity; identify critical decision points; provide appropriate levels of detail; offer feedback loops for altering certain activities if necessary and finally provide a basis for monitoring and reporting the progress of the project over time.

The DELTA chart is one such communication device which is used in Research and Development Firms in the US [1]. A form of flow chart, DELTA is an acronym for the five types of graphic symbols used in the chart, i.e. Decision box, Event box, Logic box, Time arrow and Activity box; these symbols are shown in Fig. 1. The use of each is discussed by the author in [2].

The decision box, as indicated by its name, is where the project director must choose between alternatives. Normally three alternatives are depicted by a decision box: One, based on the successful completion of preceding activities, will be a decision to go on to the next set of activities; a second will suggest that some or all of the preceding activities must be redone before work can begin on the next round of activities; a third, though a more drastic alternative, is that the decision is made to stop current activities until correcting strategies are implemented. For identification purposes the decision-maker's title (e.g. the Project Director) is written in the decision box.

For long, complex projects DELTA is superior to other network techniques through its increased flexibility (allowing for alternatives) for planning and its decision and logic elements. Such projects may also employ several organizations or persons and it is imperative that they see how they must function together to complete the project. DELTA is an excellent medium for communicating this interdependency.

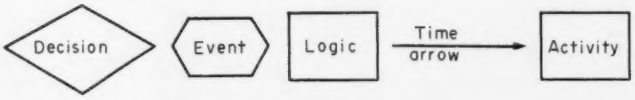


FIG. 1. *DELTA chart symbols.*

In short, for the project manager the DELTA chart provides a number of increased capabilities. DELTA makes it possible to:

1. Plan in advance the research design of the project's objectives.
2. Predict the time required to complete the project.
3. Spell out the research activities, events and decision points and determine responsibilities.
4. Provide alternatives for the research activities.
5. Measure performance against the project plan.
6. Control the project's progress through feedback.
7. Use the instrument as a communication medium for all those involved.

REFERENCES

1. WARFIELD J and HILL JD (1971) The DELTA chart: a method for R&D project portrayal. *IEEE Trans. Engng Mgmt* **EM-18** (4), 132-139.
2. HARRELL ALLEN T (1976) The DELTA chart: an aid to project management. *J. Res. Administrators.* July.

T Harrell Allen
(*July 1976*)

Speech Communication
Founders Hall 409
University of Southern California
University Park, Los Angeles
CA 90007
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Quality Control Strategy

INTRODUCTION

TO DATE, few models recommended for the determination of quality control (QC) strategies have considered stochastic variables and/or probabilistic constraints. However, the mathematical programming technique chance-constrained programming provides the methodology for incorporating these challenging aspects of functioning QC systems. Thus, several thorny difficulties faced by practitioners in the QC area no longer have to be assumed away.

There are three specific areas where such assumptions are particularly unfounded and suboptimal:

1. Almost every model developed to determine the optimal process control inspection policy makes the assumption that the process is in a state of 'statistical control';
2. Almost every model assumes that inspection accuracy remains constant throughout the entire sampling process and, indeed, at all hours of the shift;
3. Almost every model assumes that the relative risks and costs associated with the alpha and beta errors in any sampling policy are somehow balanced against each other or otherwise taken care of.

Needless to say, these assumptions greatly simplify model construction but in many instances can lead to drastically inappropriate conclusions. The first assumption establishes statistical regularity and 'guarantees' that the process average begins at standard, is reset exactly to standard with every adjustment, and does not fluctuate between samples from the process.

The second assumption ignores numerous behavioral factors operating in the quality control (QC) arena. Sound research has demonstrated that inspection accuracy deteriorates due to fatigue, boredom, differences in motivation and care exercised by individuals, and other systematic and random differences due to environmental factors (see Shein [5]).

Finally, to specify relative risks and costs of decision making errors is no easy task. However, to ignore the problem or assume it away is far from a rational or optimal approach. Frequently, the component risks and costs are somewhat easier to ascertain than the aggregate. Hence, the former approach can be successfully exploited in model formulation.

A further point should be made relative to stochastic variables. Many, if not most, models developed in the process control area are predicted on the assumption that the decision variables are deterministic or can be adequately described using expected values. In refinement and to overcome this shortcoming, chance-constrained programming allows for stochastic variables and enables the decision maker to consider entire probability distributions of feasible values rather than just expected values.

A CHANCE-CONSTRAINED PROGRAMMING APPROACH

Chance-constrained programming (CCP) was developed through the pioneering efforts of Charnes and Cooper [2, 3]. CCP optimizes an objective function subject to probabilistic constraints (i.e. constraints which can be violated a given percentage of the time) where the parameters of the model are allowed to vary according to probability distributions.

The chance-constrained programming approach is appropriate and advantageous in any quality control situation where there are

competing plans or strategies for determining when the process is in control, for determining what kind of a sampling plan is optimal for either destructive or non-destructive testing, or for other similar applications.

Because of the wide variety of situations where the approach can be applied and because the specification of various sampling plans would be highly product and firm dependent, a general chance-constrained formulation is presented which can be tailored to the specific circumstances of a given firm.

Consider that a firm is trying to formulate an overall QC program which will specify: (1) a sampling strategy, (2) a strategy to search for an assignable cause, and (3) a strategy for taking corrective action given the results obtained. Consider further that various sets of strategies have been proposed and will be denoted by the $\mathbf{x}$ vector. A general statement of a chance-constrained model would be:

$$\text{Min. } E[(\mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_3 + \mathbf{c}_4 + \mathbf{c}_5 + \mathbf{c}_6)\,\mathbf{x}]$$

Subject to:

$Pr.$ {Process is 'in control'} $\geq \alpha_1$,

$Pr.$ {An item is judged defective by the inspector when in fact it is defective} $\geq \alpha_2$,

$Pr.$ {An item is judged non-defective by the inspector when in fact it is non-defective} $\geq \alpha_3$,

$Pr.$ {An assignable cause exists and is located when one is indicated by the sampling plan} $\geq \alpha_4$,

Plus: other deterministic operating and economic constraints which are relevant to the firm or the QC function.

$$\text{Plus: } \sum_{i=1}^{n} x_i = 1 \quad x_i = \{0,1\}$$

$$i = 1,2,\ldots,n;$$

$$0 \leq \alpha_j \leq 1.0 \qquad j = 1,2,3,4,$$

where $\mathbf{c}_1$ = Cost of operating a sampling plan to determine product quality;

$\mathbf{c}_2$ = Cost of searching for an assignable cause;

$\mathbf{c}_3$ = Cost of major repairs to the equipment;

$\mathbf{c}_4$ = Cost of minor adjustments to the equipment;

$\mathbf{c}_5$ = Direct out of pocket costs of making decision errors relative to defective units;

$\mathbf{c}_6$ = Opportunity costs associated with the relative risks involved in decision errors about defective units.

The above formulation includes in the objective function all of the relevant out of pocket and opportunity costs of operating a process control system. Of course, any of these costs can be stochastic. The four probabilistic constraints are aimed at correcting the inappropriate 'in control,' behavioral, and 'relative risk' assumptions discussed in the introduction. The deterministic constraints would describe resource limitations such as manpower availability, dollar amount of expenditures, etc.

The parameter inputs to the model would be the costs and operating characteristics associated with several competing sampling plans, strategies for searching for assignable causes, and strategies for taking corrective action. Other decision variables could also be incorporated. The present decision variables x_i would be {0,1} integer variables where a value of unity would indicate the overall QC program (consisting of a sampling plan, search strategy, and corrective action strategy) that the model determined was optimal; all other programs would take on values of zero denoting rejection. Other decision variables could either be integer valued or continuous. For example, the above model could also determine the optimum number of workers to have doing the sampling (i.e. an integer valued decision variable) and the number of dollars to spend on plant-wide 'total quality control' education (a continuous decision variable).

Once a firm has specified the above aspects of the CCP model, it must be transformed into its 'deterministic equivalent' in order to arrive at a solution to the problem. The deterministic equivalent requires that two transformations be performed on the CCP formulation: (1) the expected value operator must be removed from the objective function and (2) the probability statement must be removed from the chance-constraints. Both of these modifications can be made without too much difficulty when the stochastic variables are statistically independent. See Charnes and Cooper [4] for the details of finding the deterministic equivalent under various assumptions. Frequently, the transformed problem is a linear or mixed-integer linear program which is directly solvable using existing computer-based algorithms (e.g. Balas [1]).

CONCLUSION

In summary, the new model has the following major strengths:

1. Its assumptions are much more tenable and practical than those incorporated in most existing approaches;
2. Various strategies can be evaluated on many different levels—total cost, manpower requirements, supervision necessary, behavioral considerations, effectiveness, efficiency, etc.—with an optimal overall QC program being selected for implementation, and
3. The new model takes a 'total QC system' view in which the optimal strategy for the entire QC effort is ascertained rather than the piecemeal approach taken by most existing models.

REFERENCES

1. BALAS E (1965) An additive algorithm for solving linear programs with zero-one variables. *Ops Res.* **13**(4), 517–546.
2. CHARNES A and COOPER WW (1961) *Management Models and Industrial Applications of Linear Programming*, Vols. I and II. Wiley, New York.
3. CHARNES A and COOPER WW (1959) Change-constrained programming. *Mgmt Sci.* **6**(1), 73–80.
4. CHARNES A and COOPER WW (1963) Deterministic equivalents for optimizing and satisficing under chance-constraints. *Ops Res.* **11**(1), 18–39.
5. SHEIN JB (1968) Motivational determinants of output quality. Unpublished DBA dissertation, Indiana University.

Thomas J Hindelang
(*June 1976*)
Department of Finance and Statistics
Drexel University
College of Business and Administration
Philadelphia, PA 19104
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Modeling the Product Elimination Decision

AN IMPORTANT but often neglected area associated with a firm's product policy is the elimination decision. Many firms have neither a formalized product line evaluation procedure nor the models to aid in evaluating items within the line. In fact, most deletion decisions are not part of any type of formal review program, but rather done as piecemeal approaches to crisis management [1, p. 111].

The purpose of this paper is to develop a model, incorporating the notions of expected returns, product interrelationships, and levels of risk, that can be used to support management in making decisions related to the phasing out of weak products from a product line. The approach borrows from financial portfolio theory based on the work of Markowitz [3], Sharpe [4], and Weingartner [6]. The model is evaluative in nature and designed to be part of an iterative process to review product efficiency within the line on a periodic basis. The technique gives a rank ordering of products within the line taking into account the corporate attitude toward risk. The method is unique in that it considers the expected returns of individual items while simultaneously incorporating within line product interrelationships and risk.

Vol. 4
1976

Mathematical development

The model is based on the assumption that management seeks to maximize the efficiency of its product line. In general, such an objective can be expressed, using the form of Weingartner's capital budgeting model [6], as the maximization of $E(\tilde{X}) - A \cdot \mathrm{Var}(\tilde{X})$, where $E(\tilde{X})$ and $\mathrm{Var}(\tilde{X})$ are respectively, the expected value and the variance of the future returns associated with the product line, and A is the firm's rate of substitution between a reduction in risk and an increase in return. Applying this structure to the product elimination decision, the objective function becomes:

$$\text{Maximize } Z = \sum_{k=1}^{m-1} \left[\sum_{i=1}^{m} x_{ik} E(r_i) - A \sum_{i=1}^{m} \sum_{j=1}^{m} x_{ik} x_{jk} C(r_{ij}) \right]$$

where

m = Number of products in the product line;
$E(r_i)$ = Expected return associated with product i;
$C(r_{ij})$ = Covariance between the expected returns associated with products i and j;
A = Management's attitude toward risk. (If $A = 0$, management has a neutral attitude toward risk; as A gets larger, risk becomes more important to the firm and products with low variances and covariances tend to replace those with high variances and covariances);
x_{ik} = A zero-one variable that measures the efficiency of product i relative to other products in the line.

The subscript k, in both the above objective function and constraints that follow, is used for representing the relative efficiency of the products in the product line.

Two sets of constraints must be included in the formulation of the model. The first set forces all elements of row i in matrix X to be zero once the rank of product i's relative efficiency has been established. Such constraints can be described generally as:

$$x_{ik} - x_{ip} \geq 0,$$

for $i = 1, \ldots, m$;
$k = 1, \ldots, m-2$ and $p = k+1$.

When evaluating the relative efficiency associated with an m item product line, there are $m(m-2)$ constraints of this type.

The second set of constraints relates to the $m-1$ columns of matrix X and facilitates the rank ordering of the products in the product line, i.e.

$$\sum_{i=1}^{m} x_{ik} = m - k, \quad \text{for } k = 1, \ldots, m-1.$$

Note that when m products are to be rank ordered, there are $m-1$ constraints of this variety.

Collectively both sets of constraints ensure that the products in the product line will be rank ordered by relative efficiency. Thus, if the solution obtained appears in an X matrix structured as:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

the rank ordering of the four products in terms of their relative efficiency is 2 1 3 4 (i.e. product 2 is the most efficient item in the line while product 4 is the least efficient).

It is apparent that any results obtained depend heavily on the precision of the associated estimated expected returns and variances. Any inaccuracy, while unavoidable, can be reduced by utilizing the model in conjunction with a series of different estimates. The end result of the process should reflect an accurate ranking of the products under investigation in terms of their relative efficiency. Furthermore, by varying A, one can observe the levels of risk where products change their relative efficiency with respect to one another. Such information, when coupled with management's experience, insight, and attitude toward risk, makes any decision to drop products from a line a more objective one.

Solution methodology

The presence of second degree polynomial terms in the objective function requires that a zero-one integer quadratic programming algorithm be used to calculate the values of X. Laughhunn [2] describes a method, based on the concept of implicit enumeration, that can be applied to models utilizing the conceptual framework developed by Markowitz. An alternate approach [5] takes advantage of the binary nature of the variables and derives appropriate transformations that effectively reduce the polynomial terms to equivalent linear zero-one terms. Such transformations, while adding a number of constraints (two for every second degree polynomial term), make the problem easier to solve in the sense that zero-one integer programming codes are generally more readily available than zero-one integer quadratic codes.

REFERENCES

1. KOTLER P (1965) Phasing out weak products. *Harv. Bus. Rev.* **43**(2), 107–118.
2. LAUGHHUNN DJ (1970) Quadratic binary programming with application to capital-budgeting problems. *Ops Res.* **18**(3), 454–461.
3. MARKOWITZ HW (1959) *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York.
4. SHARPE WF (1963) A simplified model for Portfolio Analysis. *Mgmt Sci.* **9**(5), 277–293.
5. WATTERS LJ (1967) Reduction of integer polynomial programming problems to zero-one linear programming problems. *Ops Res.* **15**(6), 1171–1174.
6. WEINGARTNER HM (1966) Capital budgeting of interrelated projects: survey and synthesis. *Mgmt Sci.* **12**(7), 485–516.

William B Locander
Richard W Scannell*
(June 1976)

Department of Systems & Operations Management
College of Business Administration
University of Houston
Houston, Texas 77004
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Robustness of Linear Models for Decision Making: Some Comments

THE USE of linear regression models (and linear models in general) as a basis for managerial decision making has been studied extensively by both management and behavioral scientists. In a paper I wrote some years ago on the subject [5], I presented further evidence attesting to the effectiveness of linear regression models over actual managerial decisions in a dynamic production planning context. In that paper, the issue of robustness of linear regression models to (1) different information environments, (2) weighting parameters, and (3) the functional form of the model was also discussed and supported with experimental data.

In a recent paper, Ashton [1] presents an excellent summary of the psychological research in this area, which further supports the robustness of linear models to the above three factors. I would like to comment briefly on some of Ashton's remarks, and then make some additional points that I feel are relevant to developing and selecting linear models which are designed to represent human decisions. As does Ashton, in the ensuing discussion, I will assume that the models have been specified correctly, in the sense that there are no missing or inappropriate independent or predictor variables in the model. Secondly, I will assume that the intent of the model is to *predict* rather than *describe* choices. This is an important distinction which I shall elaborate upon later.

Vol. 4
1976

ROBUSTNESS OF PREDICTIONS TO WEIGHTING PARAMETERS

In discussing the robustness of linear models with respect to weighting parameters, Ashton refers principally to the use of unit weighting (or more generally, equal weighting) models. Unit weighting models are those where the independent variables are simply weighted by $-1{\cdot}0$ or $+1{\cdot}0$, depending on the direction of correlation between each independent variable and the dependent variable (a zero weight is assigned to equivocal variables). The rationale for the use of unit weighting as well as equal and random weighting is nicely summarized by Ashton [1] and is more comprehensively formulated and discussed in Einhorn and Hogarth [2] and Wainer [7]. Simply stated, since predictions made by unit weighting models are highly correlated to predictions made by linear regression models, predictions are robust with respect to the weighting parameters of a linear model. What Ashton and many others do not expressly say is that the robustness of unit (or equal) weighting models are only guaranteed if the dependent and predictor variables are scaled in standardized form (i.e., having zero mean and unit variance) [7], or more practically speaking, have a common scale and effectively equal variances. This is not the case in the production planning problem, nor can this be achieved in any meaningful way.

To illustrate, the production and work force rules for the three period forecast horizon case in [5] were

$$P_t = a_1 + \beta_1 W_{t-1} + \beta_2[E(S_t) - I_{t-1}] + \begin{Bmatrix}\beta_3 E(S_{t+1}) + \\ \beta_4 E(S_{t+2})\end{Bmatrix} + U_1, \quad (1)$$

$$W_t = a_2 + \beta_5 W_{t-1} + \beta_6 [E(S_t) - I_{t-1}] + \begin{Bmatrix}\beta_7 E(S_{t+1}) + \\ \beta_8 E(S_{t+2})\end{Bmatrix} + U_2 \quad (2)$$

where W_t = work force level in period t; P_t = production level in period t; I_t = inventory level at end of period t and $E(S_t)$ = expected (forecasted) sales in period

t. α and β are the coefficients of the predictor variables and U the disturbance terms. The object was to choose values for the decision variables P_t and W_t such that the total incremental cost (TIC) over all decision periods was minimized, i.e.

$$\text{TIC} = \sum_{t=1}^{n} [0{\cdot}20\,(W_t - W_{t-1})^2 + 0{\cdot}025\,(P_t - kW_t)^2 + 0{\cdot}02\,(I^* - I_t)^2$$
$$\text{subject to } I_t = I_{t-1} + P_t - S_t \quad t = 1,2,\ldots,n. \tag{3}$$

where k = productivity factor in units produced per period per man, I^* = desired aggregate inventory level and S_t = actual demand (sales) in period t.

In this decision situation a unit or equal weighting model would presumably be inappropriate—standardization (or normalization) of the dependent and independent variables would make no sense since the criterion function, cost, is derived from the actual P_t and W_t decisions. We could, however, using constrained regression procedures, derive unit or equal weighting models; for example, the general production decision rule would be either

$$P_t = a_2 + [E(S_t) - I_{t-1}] + W_{t-1} + E(S_{t+1}) + E(S_{t+2}) + U_3 \tag{4}$$

or

$$P_t = a_3 + \gamma\{[E(S_t) - I_{t-1}] + W_{t-1} + E(S_{t+1}) + E(S_{t+2})\} + U_4 \tag{5}$$

respectively. Even if these models were robust (which I seriously doubt), they make little sense. Does it make sense to give equal weights to the workforce variable and sales variable, which have different dimensions? Does it make sense to equally weight sales forecasts of the immediate and distant future? Obviously not (here we begin to distinguish between predictive and descriptive models)!

In sum, while unit (and more generally, equal) weighting models are robust, the data must be capable of being transformed (i.e. standardized) to assure that this robustness property holds. This is not the case for the production planning problem and presumably for many other managerial decision situations. Hence, even for predictive purposes, unit or equal weighting schemes are not necessarily universally viable alternatives to standard regression and econometric procedures.

Ashton also uses a unit weighting argument to justify the good performance (and hence, use) of linear regression models (and linear models in general) in information environments different from the ones in which they were derived. He makes the assumption that the same known independent variables are relevant and similarly correlated with the dependent variable in both environments. For the same previous reasons, his argument, while relevant for certain problems, is not appropriate for the production planning problem in particular and to management problems in general. Hence the robustness of unit or equal weighting models to changed information environments is not assured for these types of problems.

Furthermore, the results in [5] show that even when the relevant independent variables are different in the different environments, an 'inappropriate' model can often outperform individuals. This was the case in the production planning problem, when the one period linear regression models outperformed the decision makers in the known three period forecast environments.

A relevant question then is, "How can an appropriate model be chosen in the face of a changed or changing uncertain environment?" The choice of a model under such circumstances can be approached from a decision theoretic viewpoint. To illustrate, consider Table 1 extracted from [5, Table 2)], which can be viewed as a payoff or decision matrix. Let us define the set of actions as the linear models that have been developed in various known information environments (perhaps via simulation). Let us also define the states of nature as the possible actual unknown information environments. Associated with each model-environment (action-state) pair is its cost performance (consequence). If we assign prior probabilities to each environment we can compute the expected cost of using each model and then choose the model which minimizes expected cost. For example, if we assume that the possible environments are equally likely, then the optimal model is the one developed in the medium forecast error environment.

We could also postpone the choice of a model and gather more information about the environment before deciding. The concept of expected value of perfect information (EVPI) provides a guideline for determining

TABLE 1. SENSITIVITY OF LINEAR REGRESSION MODELS (LRMs) TO CHANGES IN INFORMATION ABOUT THE DEMAND ENVIRONMENT

Actual environment: true forecast error (T = 3)	Prior probability	Forecast error environment under which LRMs derive		
		Zero	Medium	High
Zero	1/3	$10,148	15,264	16,522
Medium	1/3	4853	2938	6786
High	1/3	12,669	8762	4886
Expected cost		9223	8988*	9398

* Optimal model is LRM developed for medium forecast error environment.
Note: EVPI = [1/3(10,148) + 1/3(2938) + 1/3(4886)] − (8988) = −$2997.

whether a terminal decision should be made immediately or postponed in order to gather more information [6]. In essence, EVPI indicates the maximum amount a decision can improve by removing all uncertainty. By comparing the cost of acquiring perfect information with the EVPI, we can determine whether it is wise to consider the possibility of gathering more information and postponing the decision. Moreover, we could evaluate the robustness or sensitivity of the various models to changes in the prior probability distributions. In doing so, we can define the set of probability distributions for which each model is optimal. Albeit decision theory does not assure the use of the 'appropriate model,' it provides us with the one that will cause us the least damage.

It should be stressed that the objective in [5] was to develop *descriptive* rather than *predictive* models. The focus was thus on the interpretation of the model coefficients as measures of importance of the various predictor variables in arriving at production and work force decisions. It should be clear that this issue is somewhat different from that encountered when the purpose of the model is to simply predict decisions. In the latter case, the estimates of the individual coefficients are not of central interest—interest instead centers on the estimation of the dependent variable (hence, unit, equal, or random weighting models could be appropriate and useful). This distinction is also important in the estimation of the parameters of the model where, for example, one must be aware of the dangers of pooling data from various sources (i.e. not accounting for inter-source heterogeneity) or treating the parameters as *fixed* rather than *random* and/or *systematically changing* over time (i.e. not accounting for intra-source heterogeneity) [3, 4]. Although procedures which do not account for intra- and inter-source heterogeneity may actually lead to improved estimates of the dependent variable, the resulting coefficient estimates may not have the desired property of interpretability.

In conclusion, Ashton's paper fulfills the important and too frequently overlooked function of synthesizing research from other disciplines that may have application and use to management. While there are some dangers of extrapolating too much from one discipline to another, the benefits of such cross fertilization efforts can indeed be truly significant.

REFERENCES

1. ASHTON RH (1976) The robustness of linear models for decision making. *Omega* 4(5) 609–615.
2. EINHORN HJ and HOGARTH RM (1975) Unit weighting schemes for decision making. *Orgn. Behav. Hum. Perf.* **13**, 171–192.
3. FORD DL, MOSKOWITZ H and WITTINK DR (1976) Econometric modeling of individual and group multiattribute utility function. *Proceedings of Seventh Annual Conference on Modeling and Simulation*, Sponsored by the University of Pittsburgh, IEEE, and ISA.
4. MCCANN JM, MILLER JG and MOSKOWITZ H (1975) Modeling and testing

Vol. 4
1976

dynamic multivariate decision processes *Orgn. Behav. Hum. Perf.* **14**, 281–303.
5. Moskowitz H (1974) Regression models of behavior for managerial decision making. *Omega* **2**, 677–690.
6. Raiffa H and Schlaifer R (1961) *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University, Boston.
7. Wainer H (1976) Estimating coefficients in linear models: It don't make no nevermind. *Psychol. Bull.* **2**, 213–217.

Herbert Moskowitz
(*June 1976*)

Krannert School of Industrial Administration
Purdue University
West Lafayette
Indiana 47907
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

Fixed Costs Constraints and Realities

MANY writers, of whom Sussams [6] and Buxton [2] are examples, have described the several ways in which the techniques of operational research may be employed to good effect in the field of distribution. Their solutions to the various problems rely heavily upon the proper identification of operational costs, which most writers recognise to be difficult to achieve in practice.

Wilson [7], Christopher [3] and others have said that the answer lies in more accurate costing on the part of the accountant. They advocate marginal costing as opposed to absorption costing as the solution. But in discussing these methods, Batty [1] has said "Variable overhead costs and, where unsuitable costing methods are employed, even prime costs, cannot be traced to products with any accuracy. In all costing systems absolute accuracy is extremely difficult to achieve and, even if possible, may be too costly to justify its being obtained."

In a study [5] of the distribution function of the foods division of a major British company, it was evident that the cost data, although perfectly adequate and valuable for the purposes to which it was put by the accountants, was nevertheless insufficient for operational research analysis. The derivations of the marginal costs were acceptable to satisfy product performance guidelines for the respective management teams and were convenient to calculate: but they did not reflect the true operational costs of the distribution function.

This is by no means a criticism of the Company's foods division distribution costing methods: it serves only to highlight the fact that so many of the costs of distribution are in reality fixed costs, not variable. A company which provides its own transport and distribution service will naturally expect to charge the costs of that service to individual products. The purpose of the charging mechanism is to ensure that the distribution function continues to operate within budget. However sophisticated the charging processes may be, they do not necessarily equate with costs.

The same costs may be, of course, both fixed and variable over different time spans. In general, costs are fixed in the short term and variable in the long. Distribution costs are of this kind: short term operational costs tend to be fixed since factors such as vehicle fleet size, the number and location of depots and the product range are determined, and modified, infrequently. This leads naturally to times when capacity is greater than that which is needed, and other occasions when there is insufficient. Cyert and March [4] have concluded that the former is more common than the latter, and have expressed this excess capacity as 'organisational slack', an integral part of a company's organisation.

Company budgets are set annually with these slacks in mind, albeit subconsciously. In particular, the space available within company owned warehouses and depots exists whether it is used or not, as does, in general, the handling equipment within the depots and the manpower to operate it. Provided a company's goods are contained within the confines of its own warehousing facilities, it matters little to distribution whether they be fully utilised or not. If, however, stock is so high that outside warehousing must be employed and there has been no allowance in the budget for such an eventuality, distribution will justifiably express concern.

It was the recognition of the importance of the upper bounds of resource availability which influenced the development, design and eventual implementation of the system for the Company's distribution function. The investigation was concerned with the imbalance of levels of stock of some thousand items of food products in the company's several depots. It was a typical problem situation, requiring co-operation between marketing and production to achieve company policy objectives on stocks, with distribution acting as the buffer in the middle. Although it might have been argued that the problem was not of distribution's making, nevertheless it was incumbent upon them to smooth out the peaks and troughs of market-

Vol. 4
1976

ing and production activities. Distribution could only take what stocks it was offered by production, and place them ready for sale wherever it felt they would be most effective. Without a coordinating system, there were frequently some depots with excess stocks whilst others suffered shortages of the same items.

The problems of distribution centred upon the utilisation of company owned storage space: it was required that all the output of the production department should be accommodated, whilst at the same time satisfying customer demand. In such circumstances, from an operational standpoint, individual costs are of little relevance. It is instead an equitable allocation of space to the products, and the subsequent daily control of that allocation, which matters and upon which the solution was based.

REFERENCES

1. BATTY J (1975) *Management Accountancy*. Macdonald & Evans, London.
2. BUXTON G (1975) *Effective Marketing Logistics*. Macmillan, London.
3. CHRISTOPHER M (1971) *Total Distribution*. Gower, London.
4. CYERT RM and MARCH JG (1963) *A Behavioural Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ.
5. STAINTON RS (1971) A distribution system. Paper presented at the Conference on Operational Research, Lancaster, September 1971.
6. SUSSAMS JE (1969) *Industrial Logistics*. Gower, London.
7. WILSON RMS (1972) Financial control of PDM. *Int. J. Phys. Distrib.* 3(7).

RS Stainton
(*June 1976*)

Centre for OR Studies
Hatfield Polytechnic
PO Box 109
Hatfield
Herts.
UK

OMEGA, The Int. Jl of Mgmt Sci., Vol. 4, No. 6, 1976. Pergamon Press. Printed in Great Britain

International Register of Theses and Dissertations

The following are available at the Yvette Jacobson Library, Management Centre, University of Bradford, Emm Lane, Bradford, Yorkshire BD9 4JL, UK:

OGUNMODEDE AO (1972) Environment of Marketing Development in Nigeria. MSc Thesis, University of Bradford.

PALFREY CK (1973) The Closed Shop in Equity. MSc Thesis, University of Bradford.

PEET BW (1971) Vehicle Replacement Policy in the Yorkshire Electricity Board. MSc Thesis, University of Bradford.

PENDLEBURY TG (1971) An Assessment of the Role of Wholesales in the Pharmaceutical Industry. MSc Thesis, University of Bradford.

PLANT M (1973) Single versus Multiple Criteria as a Basis for Segmentation. MSc Thesis, University of Bradford.

PLATTS R (1973) The National Industrial Relations Court. MSc Thesis, University of Bradford.

PROCTOR GD (1972) Deliberate and Non-Deliberate Purchases in the Men's Outerwear Market: A Pilot Survey. MSc Thesis, University of Bradford.

ROSCOE J (1972) The Development of a Marketing Model for a Consumer Durable Product. MSc Thesis, University of Bradford.

SHACKLETON JK (1974) Behavioural Aspects of Accounting. MSc Thesis, University of Bradford.

SHARRATT M (1970) Timetabling at the Management Centre. MSc Thesis, University of Bradford.

SHEPHERD AP (1972) Dividend Policy: A Survey of Current Theories and Empirical Evidence Concerning its Influence on the Value of the Enterprise. MSc Thesis, University of Bradford.

SLATER AG (1970) The Organisational Structure for Marketing Information. MSc Thesis, University of Bradford.

STILL M (1971) Turnover of Asian Immigrant Labour in the Wool Textile Industry of the Bradford Region. MSc Thesis, University of Bradford.

THOMPSON CMT (1971) Industrial Marketing Planning. MSc Thesis, University of Bradford.

TINKER AM (1970) Organizational Change and Change Agents at Work. MSc Thesis, University of Bradford.

TREMLETT TR (1974) The Buro-Currency and Capital Markets. MSc Thesis, University of Bradford.

TURNER DR (1971) Board Composition in Newly Quoted UK Manufacturing Companies. MSc Thesis, University of Bradford.

TURNER J (1970) Demand Forecasting and Corporate Planning. MSc Thesis, University of Bradford.

WALTERS DW (1971) Output Budgeting: An Approach to Evaluating Market Research Cost Effectiveness. MSc Thesis, University of Bradford.

WHEELER J (1972) The Design of Personnel Information Systems for the Hotel Industry. MSc Thesis, University of Bradford.

Vol. 4
1976

WHITEHOUSE RC (1971) The Design and Installation of a Simple Production Control System in a Medium Sized Fabrication Company. MSc Thesis, University of Bradford.

WILD R (1966) Operational Research in Hospital Administration. MSc Thesis, University of Bradford.

WILLIAMS JP (1971) Changing a Wages Structure. MSc Thesis, University of Bradford.

WILLIAMS V (1973) Budgetary Control on Ships. MSc Thesis, University of Bradford.

WILSON RMS (1969) Management Control of the Marketing Function. MSc Thesis, University of Bradford.

The following are available at the Library, Cranfield Institute of Technology, Cranfield, Bedford MK43 0AL, UK:

AHMAD M (1971) Application of System Costing to a very Small Manufacturing Concern. MSc Thesis, Cranfield Institute of Technology.

ALLISON D (1971) Fluid Sensing for Linear Measurement. MSc Thesis, Cranfield Institute of Technology.

ATTARAN MA (1972) Application of Boolean Methods to Problems of Material Distribution. MSc Thesis, Cranfield Institute of Technology.

BALL WJE (1970) The Optimisation of Cutting Conditions for Peripheral End Milling. MSc Thesis, Cranfield Institute of Technology.

BARON P (1970) An Information System for Production Control. MSc Thesis, Cranfield Institute of Technology.

BARRAR PRN (1971) Quality Organisation with Particular Emphasis on Control of Costs in the Area Associated with the Quality of Design. MSc Thesis, Cranfield Institute of Technology.

BHATTACHARYA P (1972) Determination of Optimum Grinding Conditions for a Selection of Difficult Materials. MSc Thesis, Cranfield Institute of Technology.

BUCK DMC (1972) An Estimating System for Determining Machining Times in Bearing Manufacture using Classification. MSc Thesis, Cranfield Institute of Technology.

BUNN PJ (1973) An Investigation into Certain Organisational Factor Effects on Ship Maintainance Data. MSc Thesis, Cranfield Institute of Technology.

BURNETT N (1972) A Production Planning and Control Procedure for High Usage Value Items. MSc Thesis, Cranfield Institute of Technology.

BURT G (1970) The Effect of Variable Control Sensitivity on Tracking Task Performance. MSc Thesis, Cranfield Institute of Technology.

CAMDIBI FB (1973) Physiological Assessment of the Effects of Pacing and Work Content of Repetitive Industrial Operations. MSc Thesis, Cranfield Institute of Technology.

CARRASCO CLF (1973) An Investigation into the Forecasting Method Used at a Medium-Size Manufacturing Company. MSc Thesis, Cranfield Institute of Technology.

CHATZIMICHALIS A (1972) An Optimal Investment Policy for the Zeebrugge Ocean Container Terminal. MSc Thesis, Cranfield Institute of Technology.

GOULDING F (1970) An Examination of Stocking Policies in a Small Wine Company. MBA Thesis, Cranfield Institute of Technology.

GRIFFITHS RJ (1970) Relative Strength—An Indicator for Investment in the Equity Market. MBA Thesis, Cranfield Institute of Technology.

HEMMING DG (1970) The Growth of British Tourism Abroad. MBA Thesis, Cranfield Institute of Technology.

HOBBS CNG (1970) Nickel: A Strategic Study. MBA Thesis, Cranfield Institute of Technology.

JACKSON A (1973) An Investigation into the Organisation and Operation of a Precision N.C. Centre. MSc Thesis, Cranfield Institute of Technology.

JAIN RC (1973) Investigation into the Cutting Efficiency of Camm-Coated Carbide Tips. MSc Thesis, Cranfield Institute of Technology.

MORTIMER G (1971) The Role of Trading Stamps in Retail Markets. MBA Thesis, Cranfield Institute of Technology.

MORRIS JL (1971) An Analysis of the United Kingdom Market for Canned Edible Meats. MBA Thesis, Cranfield Institute of Technology.

MOSCHOVACOS G (1971) An Optimal Replacement Strategy for Civil Engineering Equipment. MBA Thesis, Cranfield Institute of Technology.

PLUMB FA (1972) Investment Appraisal of a Selected Company, BSR, Ltd. MBA Thesis, Cranfield Institute of Technology.

REID ADB (1972) The Problem of Assessment by Local Authorities of Consumer Satisfaction with their Services. MBA Thesis, Cranfield Institute of Technology.

RICKABY RJM (1972) An Investigation into Business Opportunities in Zaire on Behalf of a UK Group Trading Overseas. MBA Thesis, Cranfield Institute of Technology.

WALKER J (1972) A Survey of Multiple Regression Methods. MSc Thesis, Cranfield Institute of Technology.

WALTER AR (1972) A Survey of Operational Research Techniques Applicable to National Development Planning in Developing Countries. MSc Thesis, Cranfield Institute of Technology.

WILLIAMS DF (1973) An Investigation into the Application of a Vent Controlled Vortex Valve to the Control of a Hydrostatically Sealed Actuator. MSc Thesis, Cranfield Institute of Technology.

Vol. 4
1976

The following are available at the Lyon Playfair Library, Imperial College of Science and Technology, Exhibition Road, London SW7 2BX, UK:

RIZWI MS (1971) A Study of the Impact of Computers on Production Planning and Control. MSc Thesis, Imperial College, London.

SINGER KWQ (1971) Retailing of Wines and Spirits: A Survey of Consumer Attitudes. MSc Thesis, Imperial College, London.

SINGH KAJ (1972) Maximum Flows in Networks with Arc Gains. MSc Thesis, Imperial College, London.

SINGH MP (1972) Fixed Charge Transportation Problem. MSc Thesis, Imperial College, London.

SPENCER R (1972) An Approach to Local Government Planning. MSc Thesis, Imperial College, London.

STEVENS DO (1972) Job Evaluation in Britain. MSc Thesis, Imperial College, London.

STOTT RI (1972) The Economic Implications of a Nil-Growth Economy. MSc Thesis, Imperial College, London.

STUART JR (1971) A Study in Company Planning. MSc Thesis, Imperial College, London.

TAN PKK (1971) Forecasts and Forecasting Accuracy for Sales of Electricity. MSc Thesis, Imperial College, London.

THAPEN KK (1972) A Review of Approximate Methods in Queueing Theory and an Evaluation of a Queuing Model. MSc Thesis, Imperial College, London.

THOMAS KG (1971) Gilt-Edged Stocks: A Model for Highlighting Price and GRY Movements. MSc Thesis, Imperial College, London.

TOH TM (1972) Global Optimality in Non-Linear Programming. MSc Thesis, Imperial College, London.

TONG WSW (1972) Economic Study in the Transportation of Liquefied Natural Gas (LNG Carriers). MSc Thesis, Imperial College, London.

TOPALIAN KR (1971) Tree Search Methods and the Set Covering Problem. MSc Thesis, Imperial College, London.

TURNBULL DK (1972) Transfer Pricing of Computer Services. MSc Thesis, Imperial College, London.

WATERS CDJ (1972) The Scheduling of Technical Resources. MSc Thesis, Imperial College, London.

The following are available at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 30 Memorial Drive, Cambridge, MA 02139, USA:

ASHRAFI D (1972) The Development of Management Information or Decision Systems for a Municipality. SM Thesis, MIT Cambridge, MA.

BAKER RJ, JR (1972) Interaction Between Overhead Functions and Operating Functions in R&D Organisations. SM Thesis, MIT Cambridge, MA.

BARG HB (1973) Marketing the Commercial and Financial Benefits of the Bank Credit Card to Merchants. SM Thesis, MIT Cambridge, MA.

BLACK RO (1970) Use of Remote Computer Terminals by Managers: An Experiment. SM Thesis, MIT Cambridge, MA.

CALKIN DA (1972) Motivation for Job Performance—An Empirical Study. SM Thesis, MIT Cambridge, MA.

CHISM GW (1970) Graph-Theoretic Causal Analysis. SM Thesis, MIT Cambridge, MA.

COILLET DW (1971) The Management of Mission Oriented Research. SM Thesis, MIT Cambridge, MA.

COLEMAN RC (1972) The Marketing–Manufacturing Interface. SM Thesis, MIT Cambridge, MA.

COULOMBE RF (1973) An Evaluation of the Effects of Containerization on the Port of Boston. SM Thesis, MIT Cambridge, MA.

CREMER JC (1973) Health Maintenance Organizations in France: An Appraisal. SM Thesis, MIT Cambridge, MA.

DAVIS JD (1973) A Case Study: Foreign Steel, Domestic Steel, A Combination? SM Thesis, MIT Cambridge, MA.

DENNIS RA (1971) Model Cities: A Study of Planning and Control in a Resident-Dominated Public Agency. SM Thesis, MIT Cambridge, MA.

DOUGLAS BR (1970) A Decision Algorithm for the Resource Constrained Project Scheduling Problem. SM Thesis, MIT Cambridge, MA.

ELDRED WA (1972) Development of a Comparative Markov Model for Ship Overhaul Policies. SM Thesis, MIT Cambridge, MA.

FARRELL WM, JR (1970) Pricing Government Contracts—A Risk-Adjusted Rate of Return Analysis. SM Thesis, MIT Cambridge, MA.

GOLDBERG PL (1971) A Study of Mixed Ventures in Brazil. SM Thesis, MIT Cambridge, MA.

HARLOW AM, JR (1972) Exploring Rent Control with Urban Dynamics. SM Thesis, MIT Cambridge, MA.

HORNER KR (1972) An Experimental Graphics System for Simulations. SM Thesis, MIT Cambridge, MA.

INGERSOLL JE, JR (1973) An Examination of Relative Security Pricing as a Method of Superior Portfolio Formation. SM Thesis, MIT Cambridge, MA.

JARMOLYCH A (1973) A Computerized Interactive System for the Retrieval of Case Law. SM Thesis, MIT Cambridge, MA.

JUST J (1973) Changes in Energy Consumption 1963–1980. SM Thesis, MIT Cambridge, MA.

KAMPE WR II (1972) Multistage Production of Seasonal Goods with Long Lead Components. SM Thesis, MIT Cambridge, MA.

LAWRENCE DA (1971) Automated Courtroom Scheduling. SM Thesis, MIT Cambridge, MA.

MEMISHIAN SE (1970) Adaptation of a Consumer Response Model to the Evaluation of Political Perceptions, Attitudes, and Opinion Formation. SM Thesis, MIT Cambridge, MA.

MILLER N (1971) The Computer's Role in the Educational Process. SM Thesis, MIT Cambridge, MA.

MUDGETT JS (1972) Some Factors Influencing the Basic Labor Policies of the United States Steel Corporation. SM Thesis, MIT Cambridge, MA.

NEWTON RI (1973) A Decision-Oriented Management Information System in a Naval Shipyard. SM Thesis, MIT Cambridge, MA.

POWELL R (1972) Comparison of Techniques Using Financial Ratios as Predictors of Failure. SM Thesis, MIT Cambridge, MA.

REED CH (1971) Translation of Behavioral Theories of Management to First-Time Factory Supervision. SM Thesis, MIT Cambridge, MA.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 4
1976